Universidade de Lisboa

Faculdade de Ciências

Departamento de Biologia Vegetal

New method for the simultaneous identification of pathogenic human viruses using multiplex PCR

Adriana Cardoso Resende

Dissertação
Mestrado em Biologia Molecular e Genética

2014

.

Universidade de Lisboa
Faculdade de Ciências

Departamento de Biologia Vegetal

New method for the simultaneous identification of pathogenic human viruses using multiplex PCR

Adriana Cardoso Resende

Dissertação
Mestrado em Biologia Molecular e Genética

Orientadores: Doutor Filipe Pereira e Professora Doutora Filomena Caeiro

2014

“I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale.”

Marie Curie

Agradecimentos

No final desta longa etapa, só me resta agradecer a quem, de uma forma ou de outra, me ajudou a chegar até aqui.

Primeiramente, ao meu orientador Filipe Pereira, pela oportunidade de voltar a trabalhar com ele, pelos ensinamentos, pela disponibilidade, pela paciência que ainda mantem para mim ao fim de já uns anos de trabalho em conjunto. Por toda a fantástica orientação que me deu sem a qual este trabalho não seria possível.

Ao grupo de Genética Populacional, do IPATIMUP, por me ter acolhido e disponibilizado os meios necessários para a realização da presente dissertação.

Prof. Dra. Filomena Caeiro, minha orientadora na Faculdade de Ciências, obrigada por todos os rápidos esclarecimentos a questões que foram surgindo ao longo deste trabalho e por toda atenção e simpatia características.

Um agradecimento especial ao Hospital de Santo António no Porto, Prof. Dra. Mª Helena Ramos e Dra. Ana Constança, e ao *MRC - Universtity of Glasgow Centre for Virus Research* pela disponibilização das amostras.

Agradeço aos meus amigos, os de sempre e os que esta aventura me deu o prazer de conhecer. À Renata pelo apoio, boa disposição e presença constante. À Tatiana, ao Diogo e ao José que ao longo destes anos continuam a fazer parte da minha vida. Ao Pedro Andrade por ainda ter a paciência de me ouvir quando já está farto de vidas académicas. À Magdalena e ao Ricardo que, em cidades diferentes, foram os melhores colegas de casa que poderia ter tido. Ao Pedro Ferreira pela companhia e conversas “mesmo à biólogos” durante as curtas pausas na rotina. E, por último, mas não menos importante, às minhas duas companheiras de guerra desta etapa, que sempre estiveram disponíveis para me aturar, à Joana e à Priscila.

A ti, João, um sincero “obrigada” por… tudo! Apareceste mesmo no início desta etapa por Lisboa, marcando sempre a tua presença. Espero poder contar contigo em todas as outras que o futuro reserva.

Obrigada aos meus pais por me apoiarem incondicionalmente em todas as minhas escolhas, pela educação e pelos sacrifícios para me satisfazerem a minha necessidade de saber mais e por me transmitirem conhecimentos que nenhuma faculdade me poderia dar. À minha irmã por

me fazer rir em alturas menos boas. E, como não poderia deixar de referir, a verdadeira fonte de força e inspiração, ao meu padrinho.

Resumo

Os vírus da imunodeficiência adquirida tipo 1 (HIV-1), hepatite C (HCV) e hepatite B (HBV) constituem uma ameaça mundial para a saúde pública. Estatísticas recentes sugerem que milhões de pessoas possuem SIDA, hepatite C ou B, sendo que co-infeções entre estes três vírus são relativamente comuns devido à partilha de rotas de transmissão. Por este motivo, um melhor diagnóstico destas infeções e co-infeçoes é vital para monitorizar estas doenças infeciosas e controlar epidemias.

Um marco importante no diagnóstico clínico consistiu na implementação de testes serológicos baseados na deteção de antigénios e anticorpos antivirais. De uma forma geral, testes serológicos fazem a deteção indireta dos vírus, estando por isso sujeitos a falsos negativos, como é exemplo a elevada taxa de falsos negativos no período janela da infeção. De forma a evitar estes problemas, várias técnicas baseadas em ácidos nucleicos têm sido introduzidas para deteção direta destes vírus, com especial relevância no controlo da qualidade das doações para bancos de sangue.

A sensibilidade das técnicas baseadas em ácidos nucleicos aumentou consideravelmente com o aparecimento da reação em cadeia da polimerase (PCR), que permite a amplificação de secções do genoma viral partindo de uma quantidade pequena de amostra. De entre as várias técnicas disponíveis, as mais utilizadas são o PCR em tempo real, essencialmente utilizado na quantificação de cargas virais e monotorização da resposta às terapias antivirais, e o *multiplex* PCR, onde várias regiões são amplificadas em simultâneo permitindo a identificação viral. No entanto, a grande maioria dos métodos atualmente disponíveis baseiam-se na análise de uma única região do genoma viral (estando sujeitos a falsos negativos) ou requerem várias reações independentes, o que aumenta significativamente o custo do diagnóstico. Desta forma, é urgente o desenvolvimento de métodos que detetem simultaneamente vários vírus numa só reação.

Nesta dissertação é descrito o desenvolvimento de três PCR *multiplex* para a deteção do HIV-1, HCV e HBV e um *multiplex* PCR para a deteção simultânea destes três vírus. Após o alinhamento de várias sequências genómicas de cada vírus, 33 pares de *primers* de PCR foram desenhados para amplificação por PCR de oito regiões genómicas no HIV-1, 14 no HCV e 11 no HBV. Destas regiões, cinco foram amplificadas com sucesso em PCR *singleplex* para o HIV-1, sete para o HCV e 11 para o HBV.

De seguida, cinco regiões para o HIV-1 e HBV e seis para o HCV foram incluídas nos PCR *multiplex* individuais de forma a obter amplificações com tamanhos diferentes. Deste modo, cada vírus é identificado pela amplificação de diferentes regiões genómicas que resultam num padrão específico de bandas observáveis em eletroforese.

A eficácia dos PCR *multiplex* foi testada em 70 amostras de plasma de indivíduos residentes em Portugal e na Escócia infetados com estes vírus (32 com HIV-1, 26 com HCV, oito com HBV, três co-infeções HIV-1/HCV e uma co-infeção HIV-1/HBV). A deteção dos vírus foi concretizada pela visualização dos produtos amplificados em gel de agarose.

Os resultados comprovam que pelo menos uma região genómica foi sempre amplificada em todas as amostras. Das amostras de HIV-1, 31 % foram amplificadas para as cinco regiões, sendo representativas dos vários genótipos do HIV-1 (B, C, G e recombinantes). Metade das amostras de HCV pertencentes a diferentes genótipos (1A e 3) foram amplificadas para as seis regiões. No caso do HBV, 33 % das amostras foram amplificadas com todos os pares de primers testados. Em média, 3,31 regiões foram amplificadas no HIV-1, 4,54 regiões no HCV e 3,44 regiões no HBV. Estes testes demonstram que os PCR *multiplex* permitem a correta deteção dos vírus independentemente dos genótipos e em amostras com diferentes proveniências.

De forma a demostrar a robustez dos PCR *multiplex*, foram realizados testes de sensibilidade através de amplificações PCR com gradientes de temperatura de *annealing* e com gradientes de concentrações virais (i.e., diluições em série). Os resultados indicam que o intervalo ideal de funcionamento dos três PCR *multiplex* em termos de temperaturas de *annealing* é de 54,5 °C a 58,3 °C. No entanto, algumas regiões amplificaram em temperaturas fora deste intervalo, sugerindo que mesmo em diferentes condições podem permitir a deteçao viral. Os limites de deteção em termos de carga viral foram estimados para 343,75 cópias/mL para o HIV-1, 418,56 UI/mL para o HCV e 19,92 UI/mL para o HBV. Estes limites de deteção sugerem que os PCR *multiplex* são adequados ao diagnóstico viral durante o período janela. Foram ainda realizados alguns testes adicionais onde ficou evidente que os PCR *multiplex* são específicos e não originam amplificações na presença de material genético humano ou de outros vírus.

Por fim, foram selecionadas duas regiões de cada vírus para serem incluídas no PCR *multiplex* para a deteção simultânea de HIV-1, HCV e HBV. A escolha das regiões foi realizada tendo em consideração o sucesso de amplificação dos vários marcadores incluídos nos PCR *multiplex* individuais, bem como a conjugação de tamanhos dos produtos de PCR de forma a obter um

padrão de bandas identificadoras de cada vírus. Este *multiplex* foi testado em amostras de indivíduos co-infetados (duas HIV-1/HCV e uma HIV1/HBV) e em amostras de misturas artificiais de material genético dos diferentes vírus. Em todas as amostras foi possível a correta deteçao dos vírus presentes na mistura, com exceção de uma co-infeção HIV-1/HCV onde o HIV-1 não foi detetado, muito provavelmente devido à baixa carga viral deste vírus na amostra (41cópias/mL).

Uma das grandes vantagens dos métodos propostos nesta dissertação é a localização dos *primers* em regiões genómicas conservadas (i.e., com poucos polimorfismos), o que permite a deteção de vários genótipos. No entanto, os vários marcadores testados têm diferentes eficácias de amplificação nas amostras testadas. Este facto poderá dever-se à extrema variabilidade genética presente nestes vírus (especialmente os de RNA, HIV-1 e HCV), às baixas cargas virais de algumas amostras (especialmente em infeções com HIV-1) ou resultar de outros fatores inerentes ao próprio PCR (por exemplo, interações entre primers). No entanto, a ausência de amplificação de alguns marcadores em algumas amostras não impediu a deteçao dos diferentes vírus. Mesmo que uma ou mais regiões não sejam amplificadas, uma correta deteção dos vírus é possível com os nossos PCR *multiplex* devido ao uso de vários marcadores em simultâneo, o que constitui uma vantagem em relação aos métodos que utilizam apenas um par de *primers*.

Em conclusão, os métodos apresentados nesta dissertação são adequados para a identificação de HIV-1, HCV e HBV (separadamente ou em simultâneo) em casos de co-infeções ou quando não é adequado a utilização de testes serológicos (por exemplo, no período pré-seroconversão e em recém-nascidos). O PCR *multiplex* para a deteção de vários vírus na mesma reação facilita a análise de amostras em grande escala, ao permitir uma redução nos custos e no tempo de diagnóstico. Desta forma, estes PCR *multiplex* poderão ser usados em amostras obtidas em contexto clínico e em bancos de doação de sangue, aumentando o rigor da deteção e a rapidez na obtenção de resultados.

Abstract

Human immunodeficiency virus type 1 (HIV-1), hepatitis C virus (HCV) and hepatitis B virus (HBV) are a significant threat to public health worldwide. Several serological and nucleic acid assays are used for the correct detection of these viruses. However, most detection methods lack accuracy and are expensive and time-consuming. It is therefore of paramount importance to develop methods for the simultaneous identification of different viruses at low cost and with high accuracy.

In this thesis, one multiplex PCR for the simultaneous identification of HIV-1, HCV and HBV and three multiplex PCRs for identification of each virus are described. A total of 33 primer pairs were designed in conserved genomic regions for detection of different genotypes. Primers were initially tested by singleplex PCR and five (for HIV-1 and HBV) and six (for HCV) primers pairs were selected for final multiplex PCRs. Because the target regions were selected with a unique combination of lengths, each virus can be identified by band pattern analysis. The reproducibility of the assays was tested in 70 plasma samples from infected individuals. All samples yielded positive amplifications, with 31.3 % (HIV-1), 50 % (HCV) and 33.3 % (HBV) of the samples amplified for all target regions. The multiplex PCR for simultaneous detection of the three viruses successfully identified 83.3 % of the co-infected samples. Finally, multiplex PCRs proved to be specific and had low detection limits: 343.75 copies/mL for HIV-1, 418.56 IU/mL for HCV and 19.92 IU/mL for HBV.

The main advantages of these assays are the use of multiple targets (avoiding false-negative results), use of PCR primers designed in conserved genomic regions (increasing the specificity), use of low cost laboratory equipment and the suitability for identification of co-infections. Overall, these assays can be used for a rapid screening of samples in clinical diagnostic or blood banks.

Key words: viral identification, blood-borne viruses, multiplex PCR, co-infections

Table of Contents

Table of figures and tables

Introduction

Some of the most devastating diseases and epidemics in human history were caused by viral infections. Coupled with disease and death, viruses can also be the cause of significant chronic diseases and cancers [1]. The human immunodeficiency virus type 1 (HIV-1), hepatitis C virus (HCV) and hepatitis B virus (HBV) are among the most dangerous viruses to humans, being a significant threat to public health worldwide [2]. HIV-1 is associated with the acquired immunodeficiency syndrome (AIDS), which leaves patients open to a wide variety of other opportunistic infections. The latest statistics of the global HIV-1 epidemic show that there are 35 million people living with HIV/AIDS and that 1.5 million AIDS deaths occurred in 2013 [3]. HCV and HBV are both responsible for hepatitis and for hepatocellular carcinoma and cirrhosis. Chronic hepatitis C affects 130-150 million people worldwide [4], while more than 240 million people are infected with HBV [5]. HCV and HBV are also the leading cause of liver cancer in the world, accounting for 78 % of the reported cases [6]. A vaccine against hepatitis B has been available since 1982, but unfortunately no effective vaccines exist for HIV-1 and HCV.

There is a remarkable prevalence of co-infections among these viruses, which is related to the sharing of same transmission routes, such as blood and blood products [7]. Individuals with HCV/HBV co-infection tend to have more severe liver injuries, a high probability of liver cirrhosis and to develop hepatocellular carcinoma. The prevalence of HCV/HBV co-infections remains unknown, despite its clinical importance and the interest on its diagnosis for improved treatment. Some studies reveal that HCV/HBV co-infections are common in highly endemic areas and among subjects with a high risk of parenteral infections [8]. Moreover, it has been clearly shown that HCV and HBV infections are highly prevalent among HIV-1 infected individuals [9]. In HIV-1 infected individuals, about 10 - 12.5 % have HCV co-infection and 5 - 10 % have chronic HBV co-infection [7]. The clinical importance of HIV-1/HCV co-infection has increased over the last years since HCV-related liver disease has emerged as a major cause of deaths in HIV-1 infected patients treated with highly active anti-retroviral (HAART) [10].

The genomic structure of these three viruses varies considerably. While HIV-1 and HCV are single stranded RNA viruses, HBV is a partially double stranded DNA virus. Moreover, HIV-1 is a retrovirus and its RNA is reverse transcript in a double strand DNA and randomly integrated into cell genome (provirus), while HCV RNA serves as viral messenger RNA being directly translated. The HBV genome is considerably smaller (3.2 kbp) than those of HIV-1 (9.75 kb) and HCV (10 kb) [11].

The different genomic types are associated with different mutational rates and, consequently, with varied degrees of genetic variability. In general, mutation rates are considerably higher in RNA viruses than in DNA viruses, due to high error rates of the RNA-dependent RNA polymerases when compared with DNA-dependent DNA polymerases during genome replications [1]. The high genetic variability observed in viruses, especially RNA viruses, makes their molecular diagnosis very problematic since it is difficult to design effective methods (e.g. PCR primers or probes) to accurately target the highly polymorphic genomic regions.

The correct detection of these three pathogenic blood-borne viruses is crucial for the management of the infectious diseases. The development of sensitive serological tests has improved significantly the efficacy of viral detection. For instance, HIV-1 and HCV screening is done mostly by detection of antiviral antibodies, while HBV screening is achieved by detection of surface antigen (HBsAg) [12-14]. In addition to antibodies detection, HIV-1 can also be detected by p24 antigen tests, which allows direct detection of the virus. However, p24 antigen tests have a high rate of false-positives [15].

Serological tests are considered indirect detection methods. Therefore, residual risks of false-negatives emerge related to the pre-seroconversion window period, immunosilent carriage, infection with immunovariant viruses and, in case of HBV infection, due to occult carriage [16-18]. These diagnostic failures are serious concerns in blood donor centers, where screening must be done correctly. In fact, HIV-1, HCV and HBV are the most important infectious agents transmitted by blood transfusion [19, 20].

In order to overcome some of the serological shortcomings, a number of nucleic acid techniques (NATs) have been developed for direct detection of pathogens [21]. These molecular techniques have been used for multiple applications, namely in the direct detection of pathogens in clinical samples, genetic characterization of isolates, identification of genotype variations, determination of viral loads and detection of drug resistance mutations [22]. Molecular diagnostic is currently applied by many laboratories worldwide, using diverse techniques such as nucleic acid probes hybridization, ligase chain reaction, transcription-based amplification system and polymerase chain reaction (PCR) [22].

The advent of the PCR technique has increased significantly the sensitivity of NATs by allowing the amplification of viral nucleic acid from low quantities, while retaining a high specificity [23]. Several PCR-based systems are now available for viral detection and are increasingly applied in clinical and research virology [24-26]. The reverse transcriptase (RT)-PCR, nested PCR and

multiplex PCR are some of the most commonly used PCR assays [21]. Moreover, several techniques are also available for the analysis of the PCR amplified products, including agarose gel or capillary electrophoresis. More complex techniques are also available such as hybridization protection assay (HPA), DNA sequencing, RFLP analysis or DNA enzyme immunoassay (DEIA) [22].

However, NATs have some disadvantages, such as the high cost and the need for sophisticated laboratory equipment. By these reasons, NATs are difficult to apply on a large-scale routine screening and solutions are required to reduce screening time and costs [27]. Nevertheless, NATs have been an invaluable help in the prevention of infections, especially when accuracy is critical as in the screening of blood donations [19, 28].

In order to overcome the limitations inherent to the large-scale implementation of NATs, researchers often use a panel detection system to screen for several viruses simultaneously. In this regard, two different approaches have been used [29]. One method is based on the parallel analysis of a pool of samples in a single reaction or tube. However, some studies suggest that this approach reduces the sensitivity of the detection since the dilution of samples results in lower amounts of nucleic acid available for each individual amplification, which can lead to a high rate of false-negative results [30].

Another option involves the use of multiplex assays for detection of several targets simultaneously [29]. The simultaneous detection of several target sequences in a single reaction is a good option for saving time, effort and costs within the laboratory, without compromising results. It is also a good solution when test sample volume is limited, since various reactions can be performed simultaneously. Among the different available multiplex options, real-time PCR assays are the most used since they allow the quantification of viral loads. This method is therefore very effective to monitor disease progression and response to therapy, which is vital to the management of chronic infectious diseases. However, real-time PCR assays have a limited multiplex detection capacity due to the reduced number of fluorophores that could be combined and clearly distinguished. The available wave-lengths in existing equipment currently allows the detection of only three or four targets simultaneously [29, 31]. In order to screen for various viruses and/or target genomic regions, independent reactions in different tubes or plates have to be carried out [26]. Nevertheless, most detection assays to HIV-1, HCV and/or HBV are based on real-time PCR [32-35], with only a few using conventional PCR [36, 37]. These multiplex approaches require careful optimization of all parameters to avoid a decrease in sensitivity [38].

In general, available viral detection assays are either limited by the need to perform multiple reactions, or by the number of genomic regions used to identify each virus. Most methods targeting different genomic regions require multiple and independent reactions [26, 33]. On the other hand, methods that produce results on a single reaction are limited by the number of targets that can be simultaneously analyzed, for instance in real-time PCR. In most of these cases, a single genomic region is targeted, which makes these methods highly prone to false negative results caused by missing amplifications due to sequence variation in primer or probe binding sites [31].

As an alternative to real-time PCR, conventional multiplex PCR has been successfully applied in many areas of species identification and has been also successfully applied in molecular diagnosis. Recently, multiplex PCR was used as a valuable method for pathogenic identification of bacteria, fungi and viruses [38]. Because several targets are used for the identification, the confidence of the results is higher than when using singleplex PCR assays [39, 40]. Methods that produce results on a single reaction are highly prone to false negative results, which is less likely to occur in multiplex PCRs with several target regions for each virus [31]. Multiplex PCR is also suitable to detect viral co-infections, for instance by producing amplicons with different lengths for each virus, which results in a detectable band pattern [36].

The aim of the present dissertation is to develop individual multiplex PCR assays for the identification of HIV-1, HCV and HBV and one multiplex PCR assay for the simultaneous identification of the three viruses. The multiplex PCRs will include PCR primers designed in conserved regions of the viral genomes for an efficient detection of different genotypes. Moreover, target regions will be selected with a unique combination of lengths for each virus, allowing their unequivocal detection by fragment length analysis in agarose gel electrophoresis. The long-term objective of this work is to develop multiplex PCR assays that can be used in problematic cases, including viral co-infections, detection of infections in the pre-seroconversion window period, seronegative or immunocompromised individuals and also to resolve indeterminate serologic results. Overall, these multiplex PCR assays will be tested as useful tools for rapid clinical diagnostic and blood-banks screenings.

Material and methods

Sequence alignments

Three multiple sequence alignments with full or nearly full genomic sequences from HIV-1, HCV and HBV were used. The ‘HIV-1/SIVcpz compendium 2012’ alignment was downloaded from the Los Alamos public database [41]. This alignment corresponds to a carefully chosen subset of sequences printed in the HIV sequence compendium [42] and was built using HMMER [43] and edited manually to keep the reading frames intact. The HIV-1 alignment includes 170 genomic sequences representing the HIV-1 M group subtypes A (n=16), B (n=36) C (n=21), D (n=11), F (n=11), G (n=9), H (n=4), J (n=3), K (n=2) and recombinant forms (n=57). The HIV-1 reference sequence with GenBank accession number K03455.1 was also included in the alignment.

The HCV sequence alignment was downloaded from the HCV database project [44]. We used the ‘HCV Web alignment 2008’ with 471 complete genome sequences, which is a manually edited alignment including only one sequence from patient. Moreover, very similar or epidemiologically related groups are only represented in the alignment by one sequence. All HCV genotypes are represented in the alignment, namely genotypes 1 (n=378), 2 (n=29), 3 (n=7), 4 (n=11), 5 (n=3), 6 (n=43). The HCV reference sample with GenBank accession number NC_004102.1 was used.

Regarding HBV, no curated sequence alignments were available in public databases. Therefore, we built a multiple sequence alignment with complete HBV genome sequences retrieved from the NCBI Entrez Nucleotide database [45] representing distinct genotypes sampled across different world regions. We selected 386 sequences distributed by genotypes A (n=37), B (n=16), C (n=120), D (n=100), E (n=53), F (n=35), G (n=3) and H (n=22), according to the classification provided by Harrison et al [46]. The HBV reference sequence with GenBank accession number EU054331.1 was used. The sequences were aligned using the default parameters of the Muscle software [47]. All sequence alignment can be found in the IGVID database associated with this work [48].

Selection of target regions and PCR primer design

Target regions were selected to be included in the multiplex PCRs using the sequence alignments and considering the following criteria: 1) each region had to have a conserved length across the multiple sequence alignments (i.e., low intra-virus diversity in sequence length); 2) the different target regions of each virus had to be of different lengths to allow their combination in a multiplex PCR and 3) the target regions had to be bounded by conserved regions of 17 to 26 nucleotides to allow the design of PCR primers.

In order to design PCR primers that specifically delimit the selected target genomic regions in a multiplex PCR, several criteria were taking into account: 1) the potential primers were designed to avoid variable positions in the last fivA volume of 460 µL of supernatant was removed and the remaining 140 µL of pellet was used as input for the viral RNA extraction protocol. e 3' end positions; 2) primers were designed with a predicted melting temperature between 57 to 62 °C, estimated in the OligoCalc website [49]; 3) primers were tested in the AutoDimer v1 program [50] to avoid primer-dimer and hairpin interactions and 4) pairs of primers were compared with the human and viral genomes using the NCBI Basic Local Alignment Search Tool [51] to avoid nonspecific amplifications.

The following nomenclature was implemented for target regions and PCR primers: *'Virus name_amplicon length in reference sequence*' to designate the target regions. For example, HBV_101 indicates a target region for HBV with 101 nucleotides (nt). The only exception for this nomenclature regards HIV-1_LTR, which indicates a HIV-1 target region in the long terminal repeat (LTR) region and it does not indicate the amplicon length. The nomenclature *'Virus name _ primer position in reference sequence_F or R'* was used to designate the name of PCR primers. For instance, HCV_16F indicates a forward PCR primer for HCV starting at position 16 of the reference genome sequence.

Plasma samples

A total of 70 plasma samples were tested in this work, including 43 from infected individuals living in Scotland (24 infected with HIV-1 and 19 with HCV) and 27 from individuals living in Portugal (eight infected with HIV-1, seven with HCV, eight with HBV, three with HIV-1/HCV co-infections and one with HIV-1/HBV co-infection). The infectious state of the individuals was previously determined by immunological methods and the viral load quantified by real time PCR

in the Medical Research Council, Centre for Virus Research, University of Glasgow, Scotland and Santo António Hospital, Porto, Portugal.

The 32 HIV-1 samples are from the most frequent group M, namely from subtypes B (n=11), C (n=10), A1 (n=2), G (n=2) and complex/recombinants (n=5). The subtype of two samples is unknown. The HIV-1 viral loads in the samples ranged from 111 to 544,436 copies/mL. The 26 HCV samples represent genotypes 3 (n=2), 1A (n=9), 4 (n=1) or have an unknown genotype (n=4) and have viral loads ranging from 2,670 to 11,000,000 IU/mL. The eight HBV samples represent genotype A (n=2), D (n=3) or have an unknown genotype (n=4). The HBV viral loads range from 20 to 170,000,000 IU/mL. We also used samples from individuals with co-infections, namely one with HIV-1/HBV and three with HIV-1/HCV. Table 1 of supplementary material provides detailed information for all samples.

The samples were aliquoted and stored at -80 °C and were only frozen and thawed twice prior to testing. Ethical permission for using all samples in the present study was obtained in each country through the appropriate committees.

RNA isolation and cDNA synthesis

RNA was extracted from 600 µL of HIV-1, HCV and HIV-1/HCV co-infected plasma specimens using the QIAamp Viral RNA Mini Assay (Qiagen, Valencia, CA). Virions were initially concentrated from the plasma by centrifugation at full speed (16,873 g) for three hours. A volume of 460 µL of supernatant was removed and the remaining 140 µL of pellet was used as input for the viral RNA extraction protocol. Extracted RNA was eluted with 60 µL of AVE buffer provided in the extraction kit, aliquoted and stored at -80 ºC until use.

For HBV and HIV-1/HBV co-infected plasma specimens, DNA and RNA were extracted from 200 µL initial volume using PureLink®Viral RNA/DNA Kits (Invitrogen Life Technologies, Carlsbad, CA). The extracted DNA and RNA was eluted with 50 µL of sterile RNase-free water provided in the extraction kit, aliquoted and stored at -20 ºC (HBV samples) and -80 ºC (HIV-1/HBV co-infection) until use.

Reverse transcription of HIV-1 and HCV samples was performed using the SuperScript™ III Reverse Transcriptase (Invitrogen Life Technologies) and random hexamer primers. Briefly, 10 µL of HCV RNA or 5 µL of HIV-1 RNA were used in each reaction. Extracted RNA was reverse

transcribed in a total volume of 20 µL with 1 µL of dNTPs at 10 mM, 1 µL of random hexamer primers (50 µM) and 2 µL (in case of HCV) or 5 µL (in case of HIV-1) of water DNase, RNase- and protease-free (5 Prime, Germany). The RNA, dNTPs, random hexamer primers and water were first incubated at 65 ºC for 5 min and 4 ºC for 5 min. Then, 4 µL of 5X First-Strand Buffer, 1 µL of DTT (0.1 M), 0.5 µL of RNaseOUT and 0.5 µL of SuperScript™ III RT (Invitrogen Life Technologies) were incubated at 50 ºC for 60 min and 70 ºC for 15 min.

Singleplex and multiplex PCR amplifications

Initially, two samples of each virus (1IV001 and 1IV003 for HIV-1, CV001 and CV002 for HCV and BV013 and BV839 for HBV) were used to test all primers in singleplex reactions. The PCR was prepared in a total volume of 10 µL as follows: 5 µL of Multiplex PCR Master Mix (Qiagen, Valencia, CA), 1 µL of forward primer (2 µM), 1 µL of reverse primer (2 µM), 2 µL of water DNase, RNase- and protease-free (5 Prime) and 1 µL of viral cDNA/DNA. The thermocycling conditions were: initial step at 95 ºC for 15 min; 40 cycles for HIV-1 and 35 cycles for HCV and HBV at 94 ºC for 30 s, 55 ºC for 1 min 30 s, 72 ºC for 1min; with a final extension at 72 ºC for 10 min. PCRs were performed in a 2720 ThermoCycler (Applied Biosystems, Life Technologies, United Kingdom).

The selected primers were then combined in the same proportion for multiplex PCRs (three individual multiplex PCRs for each virus and one multiplex PCR for simultaneously detection of the three viruses). The multiplex PCRs were performed using the same conditions used in singleplex reactions. In the case of some HIV-1 samples with weak amplifications due to low viral loads, a second PCR with additional 20 cycles was performed using the PCR products (1 µL) from the first PCR as template.

We tested the sensitivity of the individual multiplex PCRs using gradients of primer annealing temperatures between 48 °C to 65 °C with one sample from each virus (1IV135 for HIV-1, CV1A-6 for HCV and BV151 for HBV). A second sensitivity test was performed with serial dilutions of viral concentrations. In this case, a sample of each virus of known viral load (sample 1IV001 with 44,000 copies/mL, CV1A-6 with 107,152 IU/mL and BV151 with 81,600 IU/mL) was diluted by a factor of two. The specificity tests were performed using samples from the three viruses and human DNA from an uninfected individual.

We validated the final multiplex PCR using co-infections samples and artificial co-infections made by mixing cDNA/DNA from different samples. The artificial co-infection number 1 is a mixture of 1IV268, CV753 and BV839 genetic material, while the artificial co-infection number 2 is a mixture of 1IV400, CV679 and BV646. PCRs were performed using the same conditions as previously described.

Electrophoretic separation and sequencing

The PCR amplified products were separated by electrophoresis on a 2 % agarose gel containing 1 µL of GelRed (Biotium, USA) at 140 V during 50 minutes. A PCR product volume between 3 to 5 µL was carried out in agarose gels. The fragments were visualized under UV light in a Molecular Imager® ChemiDocTM XRS Imaging System (Bio Rad, USA) with Quantity One software v4.6.9 (Bio Rad).

The two samples from each virus used to test the PCR primers by singleplex PCR were sequenced in both directions with the same primers used for PCR in order to confirm the target sequence. PCR products (1.5 µL) were purified with ExoSAP-IT® (USB, Affimetrix, USA) according to the manufacturer's recommendations. Sequencing of purified samples (2.5 µL) was performed as follows: 5 µL reactions mixtures were prepared by combining 2 µL of Big Dye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) with 0.5 µL of primer (2 µM). Thermal cycler conditions were: 96 °C for 2 min, 35 cycles at 96 °C for 15 s, 50 °C for 9 s and 60 °C for 2 min and one final hold at 60 °C for 10 min. Sequencing reaction products were purified using Sephadex™ G-50 Fine DNA Grade columns (GE Healthcare, United Kingdom) according to the manufacturer's recommendations. Purified samples were added to 12 µL of HI-DI formamide (Applichem, Germany). Sequencing was performed in a Genetic Analyzer 3130xl sequencer (Applied Biosystems), according to the manufacturer's recommendations. Sequence analysis was performed using Sequencing Analysis software v5.2 (Applied Biosystems) and Geneious version 5.4 created by Biomatters [52].

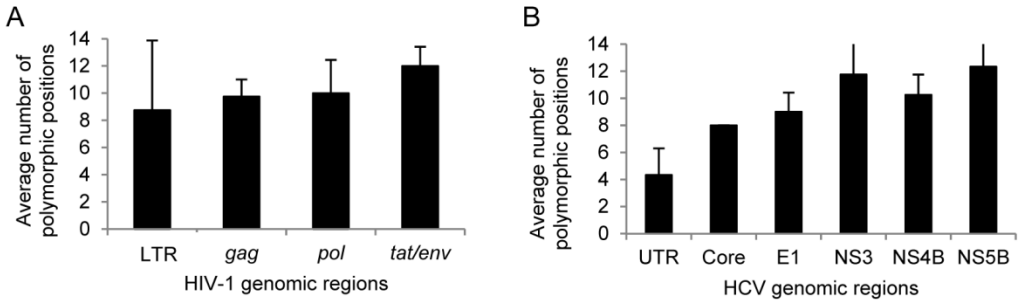
Results

Selection of PCR primers and target regions

Conserved regions were primarily identified using the identity option of the Geneious software version 5.4 [52]. A total of 16 conserved regions for HIV-1, 26 for HCV and 20 for HBV were selected. The conserved regions were chosen as potential primer-binding sites with 17 to 26 nt (Table S2) and were distributed throughout the viral genomes (Fig. S1).

In the case of HIV-1, the average number of polymorphic positions in conserved regions was 9.89 nt, with a minimum of 4 nt in HIV1_496R (with a total length of 18 nt) and a maximum of 16 nt for HIV1_9581R (with a total length of 18 nt). In the case of HCV, the average number of polymorphic positions was 9.61 nt, with a minimum value of 3 nt in HCV_16F and HCV_250F (with a total length of 17 and 20 nt, respectively) and a maximum value of 16 nt in HCV_3661R and HCV_8810R (with a total length of 23 and 20 nt, respectively). Despite the high number of sequences (n=387) in the HBV alignment, the average number of polymorphic positions in conserved regions was lower (average = 6.91 nt) than in the other viruses. The HBV presented a minimum of two polymorphic positions in HBV_1895R (with a total length of 19 nt) and a maximum of 10 nt in HBV_970F and HBV_2809F (with a total length of 23 and 22 nt, respectively), as described in table S2.

We have avoided selecting primers with polymorphisms in the last five bases at the 3' end in order to increase the primers specificity. Overall, 53 out of 66 primers were designed to have two or less polymorphic positions in the last five bases. It is interesting to notice that primers located in certain genomic regions were highly conserved, such as the HIV-1 LTR and the HCV 5' untranslated region, 5'UTR (Fig. 1A and B).



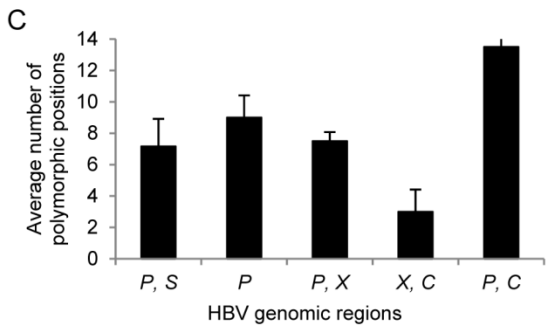


Fig. 1: Average number of polymorphic positions on PCR primers located in (A) HIV-1, (B) HCV and (C) HBV genomic regions. The polymorphisms were determined in the multiple sequence alignments made for each virus. The whiskers represent standard deviation of the mean.

In fact, the LTR and 5'UTR are often used for the design of PCR primers, possibly by being more conserved [53, 54]. However, some sequences included in the alignments were incomplete for those peripheral regions, which may help to explain the apparent high conservation, i.e., less polymorphisms are expected with less represented sequences (Table S2). The lowest average number of polymorphic positions in HBV was found in *X* and *C*, although only two primers were designed in that region.

The target regions were also selected to define genomic segments conserved in length across the sequence alignment, meaning that all viruses would yield amplicons with a similar length. This aspect is important if the length of the region is to be used for identification purposes. The standard deviation of the mean length of the target regions varies between 0 nt (i.e., no length variation) observed in regions from the three viruses to the maximum values of 0.4 nt in HIV-1 (HIV_LTR, HIV1_169 and HIV1_434), 4 nt in HCV (HCV_320) and 15.1 nt in HBV (HBV_184). A high standard deviation (7.4 nt) was observed in the marker HBV_269. The variations in the range of 0.4 to 7.4 nt are smaller enough not to cause visible changes in the mobility of bands in agarose gel electrophoresis. However, the length variation of 15.1 nt in the HBV_184 marker may cause visible changes in the separation of bands, requiring caution if used for virus identification in multiplex PCR. Interestingly, this difference results from a 33 nt deletion common to all HBV genotype D viruses (Fig. 2).

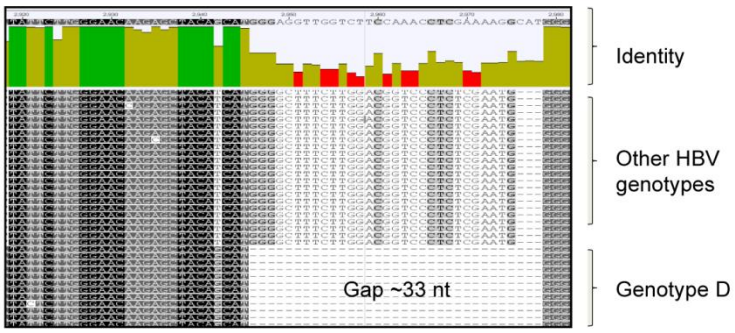


Fig. 2: Partial HBV sequence alignment. The region with a 33 nt gap is observed only in genotype D viral sequences.

The PCR primers were located throughout the entire genome of the three viruses, showing that conserved regions suitable for primer design can be found in different coding and non-coding regions (Fig. S1). Moreover, we were also able to design PCR primers in all genomic regions with the required parameters for efficient primer binding, such as a 17 to 26 nucleotide length and an expected melting temperature between 57.5 and 62.9 °C (Table S2).

Target validation by singleplex PCR

The selected targets regions (eight for HIV-1, 14 for HCV and 11 for HBV) were amplified by singleplex PCR to test the amplification efficiency. Two plasma samples of infected individual with each virus were used: 1IV001 and 1IV003 for HIV-1, CV001 and CV002 for HCV and BV013 and BV839 for HBV. The two HBV samples correspond to different genotypes (A for BV013 and D for BV839), while no genotype information was available for the other four samples.

From the eight HIV-1 target regions, five were successfully amplified in both samples (Fig. 3A). The intensity of the amplified products was lower in sample 1IV003 than in sample 1IV001 probably due to the low viral load of the former (15,400 copies/mL for 1IV003 and 44,000 copies/mL for 1IV001). This difference was particularly marked in the fragment HIV1_108, although the amplification in 1IV003 is still detectable. A total of seven regions from the 14 selected in HCV were amplified in both samples with similar intensity (Fig. 3B). The HCV targets HCV_88 and HCV_100 were of difficult distinction in agarose gel electrophoresis due to their similar sizes (88bp and 100bp). The 11 HBV target regions were successfully amplified in both samples (Fig. 3C). The 33 nt deletion common to all HBV genotype D (Fig. 2) is clearly visible in

the target region HBV_184, with two bands of predicted lengths of 184 bp (sample BV013, genotype A) and 151 bp (sample BV839, genotype D). It is interesting to notice that the PCR primer pair HBV_2809F and HBV_2992R can be used to discriminate HBV genotype D viruses from other genotypes. In any case, this target is not suitable for inclusion in the multiplex PCR to detect HBV due to the length variation that generates bands with different sizes that can confound the interpretation of results.

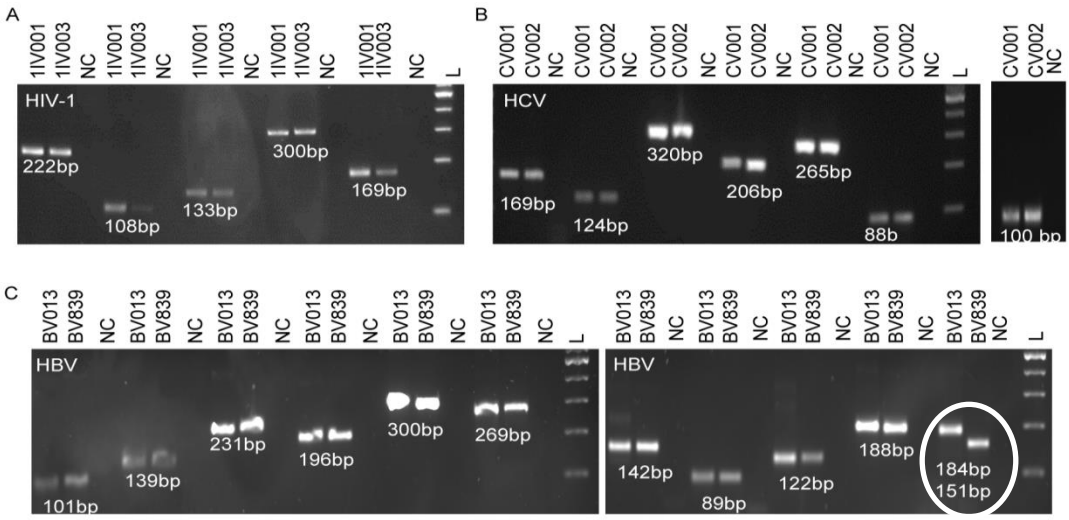


Fig. 3: Singleplex PCR amplifications of 5 HIV-1 (A), 7 HCV (B) and 11 HBV (C) target genomic regions. Only the target regions with positive amplifications are shown. Two samples for each virus were used to test all designed PCR primer pairs. NC - negative control; L -100 bp DNA ladder.

Overall, the observed band lengths on agarose gels are in agreement with the expected sizes predicted in the multiple sequence alignments. The separation of amplified fragments by agarose gel allows for a good discrimination of target regions, with exception of HCV_88 and HCV_100 targets whose length difference is small (Fig. 3B).

The sequencing of all products amplified by singleplex PCRs confirmed that all target regions were being correctly amplified (data not show).

Target validation by multiplex PCR and reproducibility analysis

The five HIV-1 target regions successfully amplified by singleplex reaction were all included in a multiplex PCR for specific detection of HIV-1 by combining all primers in equal proportions. From the seven HCV target regions, only the target HCV_100 was excluded because of the

proximity in length to the fragment HCV_88. We selected five HBV fragments to be integrated in the final HBV multiplex PCR (Fig. 4).

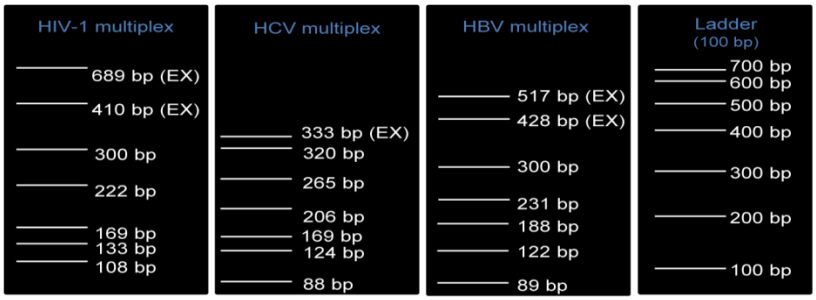
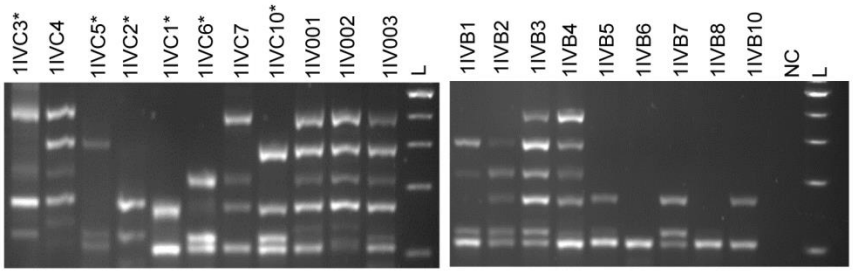


Fig. 4: Expected band pattern resulting from the multiplex PCRs for HIV-1, HCV and HBV. The multiplex PCRs were designed to originate 5 HIV-1, 6 HCV and 5 HBV amplicons, However, the additional bands highlighted with 'EX' are those that can result from the amplification with primers from different target regions.

primers HIV1_1292F and HIV1_1609R (410 bp) and HIV1_4377F and HIV1_5062R (689 bp). These extra bands did not interfere with the expected profile. In total, the HIV-1 multiplex PCR had the expected five amplified products plus two bands resulting from primers of different pairs, the HCV had the six expected amplicons and one extra from cross-amplifications and HBV had five amplicons and two extra amplified products (Fig. 4).

The reproducibility of the assay was assessed by testing multiple cases of individuals affected with each one of the virus. Firstly, 32 samples of HIV-1 infected individuals were typed using the HIV-1 multiplex PCR. At least one of the five target region was amplified in all HIV-1 samples (Fig. 5). Each of the target region was amplified in at least 50 % of the samples (Fig. 6A).



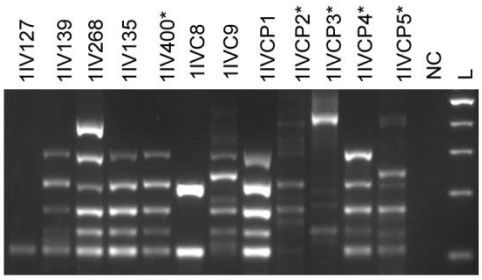
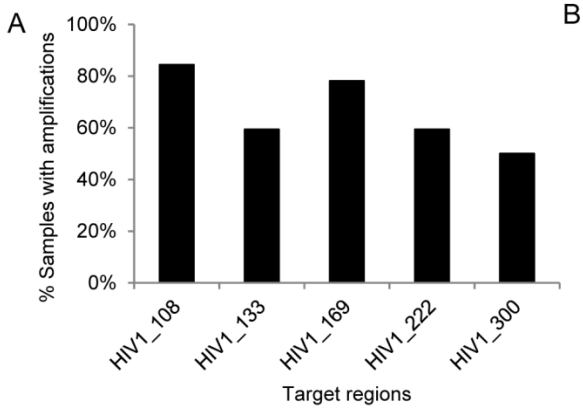


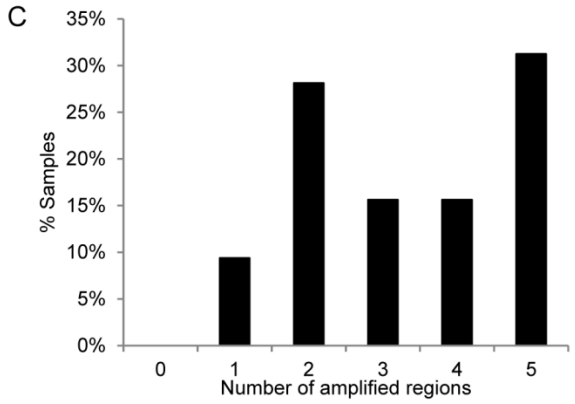
Fig. 5: Screening of samples with the HIV-1 multiplex PCR. The agarose gels show the amplified products (108, 133, 169, 222 and 300 bp) obtained in 32 HIV-1 samples.* - samples with a second PCR (20 cycles); NC - negative control; L - 100 bp DNA ladder.

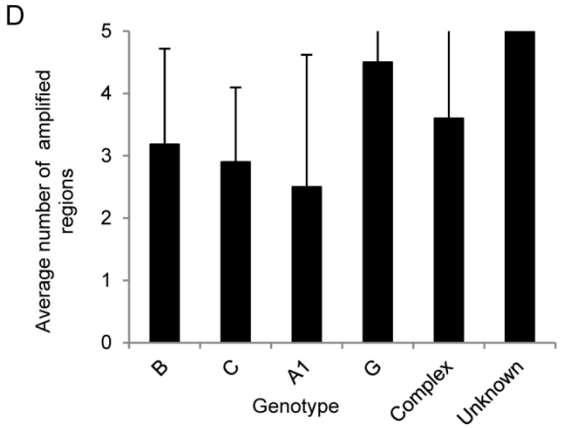


B

Frequencies of amplified target regions per genotype

Genotype	HIV1_108	HIV1_133	HIV1_169	HIV1_222	HIV1_300
B (*n=11*)	1	0.55	0.73	0.45	0.45
C (*n=10*)	1	0.86	1	0.71	0.57
A1 (*n=2*)	1	0	0.50	0.50	0.50
G (*n=2*)	1	0.50	1	1	1
Complex (*n=5*)	0.60	0.80	1	0.80	0.40
Unknown (*n=2*)	1	1	1	1	1





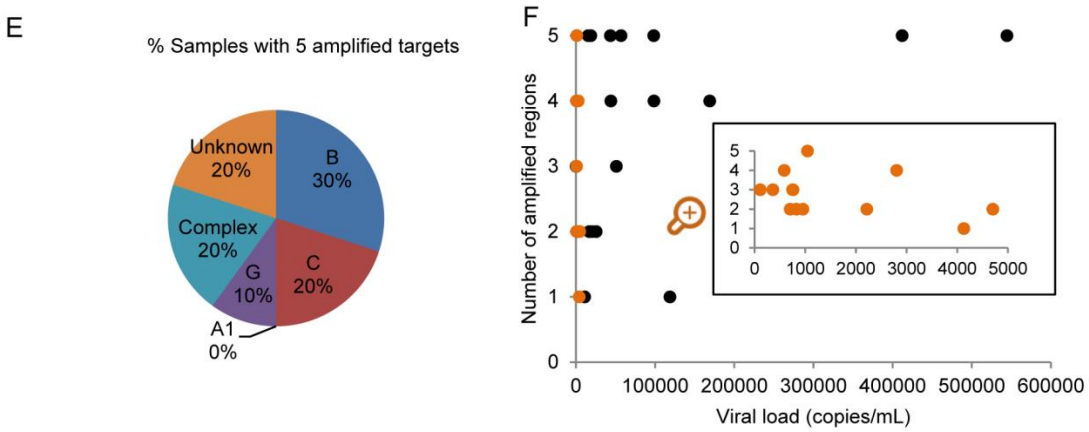


Fig. 6: Descriptive results of the screening of 32 samples with the HIV-1 multiplex PCR. (A) Percentage of samples with positive amplifications for the 5 HIV-1 target regions. (B) Frequencies of positive amplifications for the 5 HIV-1 target regions considering the different HIV-1 genotype (B, C, A1, G, complex and unknown). (C) Distribution of samples according to the number of amplified target regions. (D) Average number and standard deviation of amplified target regions per HIV-1 genotype. (E) Genotype distribution of samples amplified for all target regions. (F) Distribution of the number of amplified regions considering the viral load.

The target region HIV1_108 was the one with more positive amplifications (84.38% of the samples), while HIV1_300 had the lowest amplification success (50% of the samples), as shown in Fig. 6A. The smallest amplicons (e.g., HIV1_108) were amplified more often than the largest regions (e.g., HIV1_300).

When considering the amplification success per genotype, all HIV-1 makers except HIV_133 were amplified in samples from all genotypes (Fig. 5B). The lack of HIV1_133 amplification in genotype A1 may be influenced by the reduced sample size of this genotype (n=2), suggesting that additional samples should be tested (Fig. 6B).

About 31.3 % of the HIV-1 samples had amplifications in the five regions, while only 9 % of samples had a single amplification (Fig. 6C). On average, at least 2.5 target regions were amplified in HIV-1 genotypes (Fig. 6D). The majority of samples with the five amplified targets are from genotype B (Fig. 6E). The genotype G samples had more amplified target regions on average (4.5) than the others, when not considering those samples whose genotype is unknown (Fig. 6D). Nevertheless, the low sample size (n=2) of genotype G prevents further conclusions

The data suggests that there might be a correlation between the viral load values and the number of amplified regions. The two samples with the highest viral loads had all regions amplified, while most samples with a single amplified region have less than 500 copies/mL (Fig. 6F). However, some samples with low viral loads yield amplifications for all target regions,

suggesting that other factors might explain the different amplification efficiencies across samples.

The reproducibility of the multiplex PCR assay for HCV was assessed by testing 26 samples of infected individuals. In all cases, at least one HCV marker was amplified (Fig. 7).

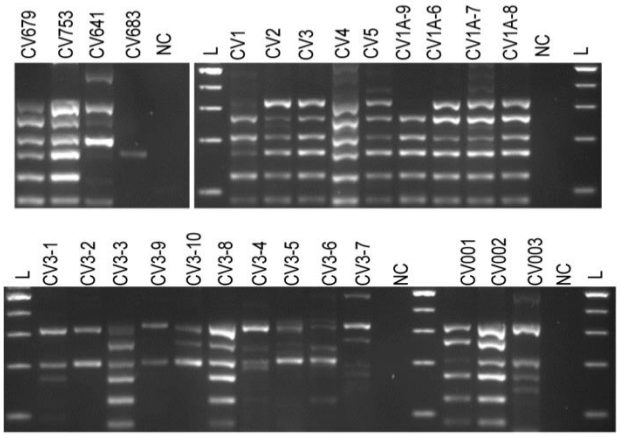
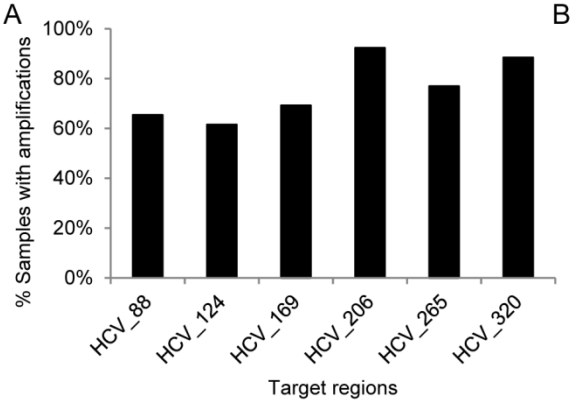


Fig. 7: Screening of samples with the HCV multiplex PCR. The agarose gels show the amplified products (88, 124, 169, 206, 265 and 320 bp) obtained in 26 HCV samples. NC - negative control; L - 100 bp DNA ladder.

Each target region was amplified in at least 62 % of the samples. The marker HCV_206 was the one with the highest amplification efficacy (92.31% of the samples), while marker HCV_124 was amplified in 61.54 % of the samples (Fig. 8A).

A



B

Frequencies of amplified target regions per genotype

Genotype	HCV_88	HCV_124	HCV_169	HCV_206	HCV_265	HCV_320
3 (*n=12*)	0.33	0.25	0.33	0.92	0.58	1
1A (*n=9*)	1	1	1	1	1	0.78
4 (*n=1*)	0	0	1	0	0	0
Unknown (*n=4*)	1	1	1	1	1	1

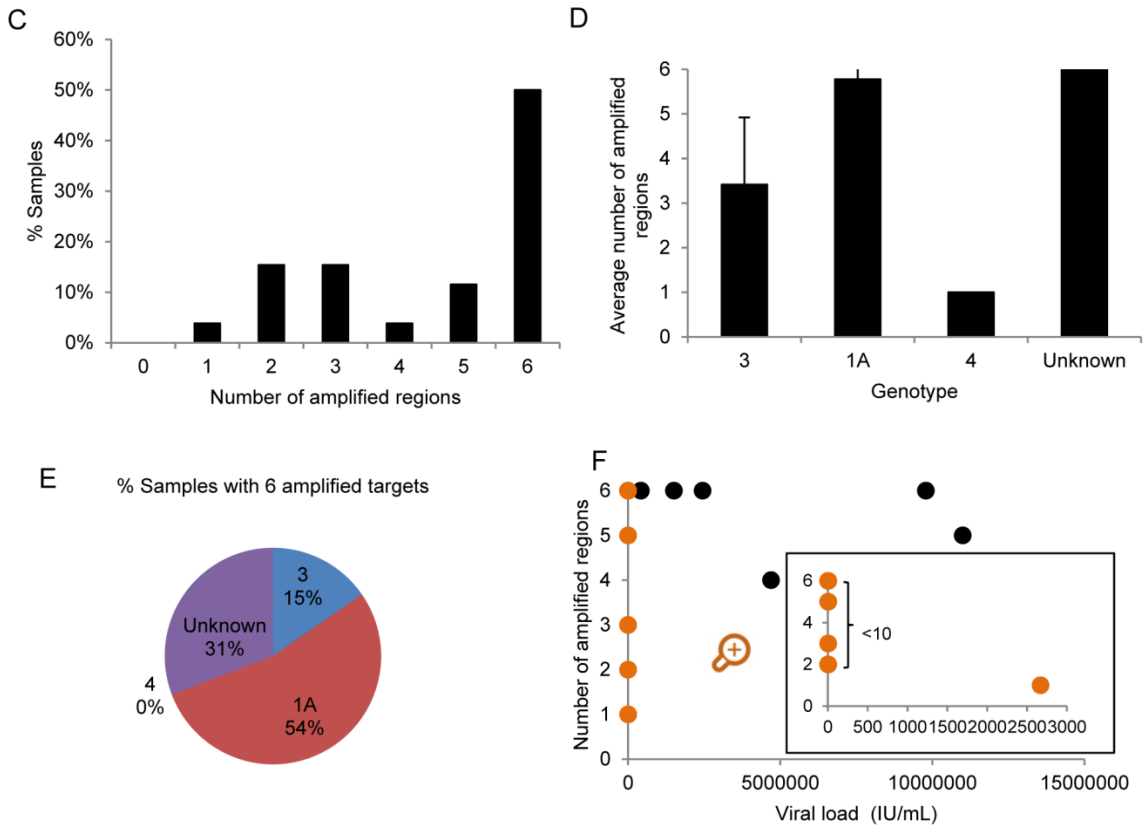


Fig. 8: Descriptive results of the screening of 26 samples with the HCV multiplex PCR. (A) Percentage of samples with positive amplifications for the 6 HCV target regions. (B) Frequencies of positive amplifications for the 6 HCV target regions considering the different HCV genotype (3, 1A, 4 and unknown). (C) Distribution of samples according to the number of amplified target regions. (D) Average number and standard deviation of amplified target regions per HCV genotype. (E) Genotype distribution of samples amplified for all target regions. (F) Distribution of the number of amplified regions considering the viral load.

In the case of HCV, the largest target regions had the highest amplification success (markers HCV_206 and HCV_320 with an amplification efficiency of 92.31 % and 88.46 %, respectively). When taking into account the genotypes, genotype 1A samples were amplified for the six markers (Fig. 8B). All samples whose genotype is unknown had the six target regions amplified. The genotype 3 samples had the highest amplification frequency in the largest markers, HCV_206, HCV_265 and HCV_320. The single sample of genotype 4 was only amplified for marker HCV_169 (Fig. 8B).

Half of the HCV samples were amplified for all target regions, while only 4 % of the samples (n=1) had a single amplified region (Fig. 8C). The average number of amplified in genotypes 1A samples was 5.78 ± 0.44 regions and in in genotype 3 was 3.42 ± 1.51 (Fig. 8D). Overall, the samples with the highest number of amplified regions were found in all genotypes, with the

exception of genotype 1A (Fig. 8E). However, additional genotype 1A samples (only one was tested in this work) are necessary to confirm these results.

Our results suggest that samples with the highest viral loads tend to have more amplified regions, although some samples with low viral loads also had six amplified regions (Fig. 8F).

Although the low number of HBV samples (n=9) used to test the reproducibility of HBV multiplex PCR requires cautious interpretation, we found that all tested samples yielded positive amplifications (Fig. 9).

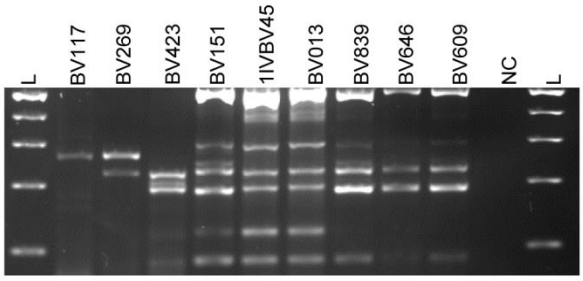
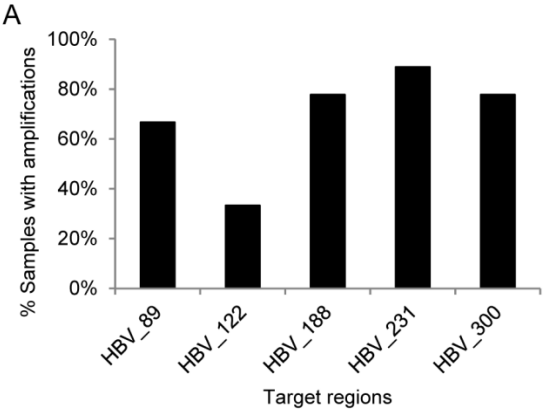


Fig. 9: Screening of samples with the HBV multiplex PCR. The agarose gels show the amplified products (89, 122, 188, 231 and 30 bp) obtained in 9 HBV samples. NC - negative control; L - 100 bp DNA ladder.

by HBV_188 (188 bp) and HBV_300 (300 bp), amplified in 77.78 % of the samples. The marker HBV_122 (122 bp) had the lowest amplification success, being only amplified in 33.3 % of the cases (Fig. 10A).



B

Frequencies of amplified target regions per genotype

Genotype	HBV_89	HBV_122	HBV_188	HBV_231	HBV_300
A (*n*=2)	1	1	1	1	1
D (*n*=3)	1	0	1	1	0.67
Unknown (*n*=4)	0.25	0.25	0.50	0.75	0.75

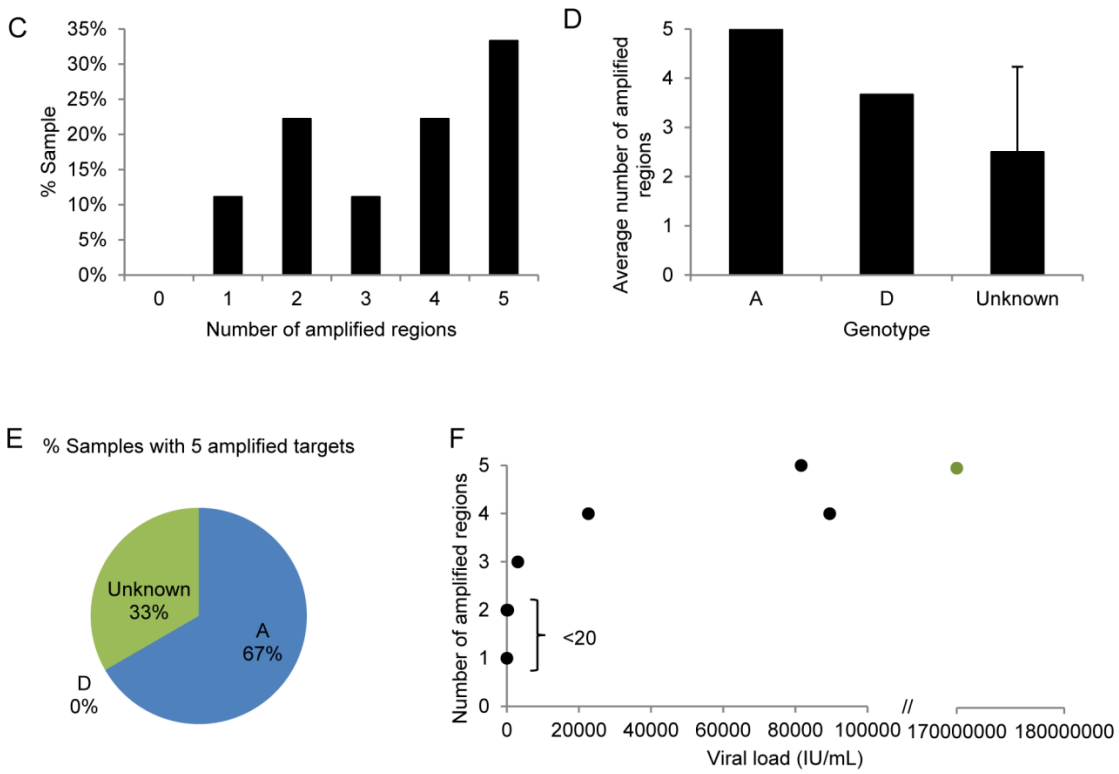


Fig. 10: Descriptive results of the screening of 9 samples with the HBV multiplex PCR. (A) Percentage of samples with positive amplifications for the 5 HBV target regions. (B) Frequencies of positive amplifications for the 5 HBV target regions considering the different HBV genotype (A, D and unknown). (C) Distribution of samples according to the number of amplified target regions. (D) Average number and standard deviations of amplified target regions per HBV genotype. (E) Genotype distribution of samples amplified for all target regions. (F) Distribution of the number of amplified regions considering the viral load.

The HBV genotype A samples were amplified for all target regions, while genotype D samples ware amplified for four target regions, missing the amplification of HBV_122 (Fig. 10B). Thus, 33.3 % of the samples had the maximum number of amplifications (Fig. 10C), being from genotype A or unknown (Fig. 10E). The average number of amplified regions in genotype D was 3.67 ± 0 (Fig. 10D).

There is a clear tendency for samples with the highest viral loads (>22,600 IU/mL) to have a better amplification success (four or five amplified regions), while samples with the lowest viral loads (< 20 IU/mL) only amplify one or two regions (Fig. 10F).

Sensitivity and specificity of the multiplex PCR assays

The sensitivity of the multiplex PCR assays was tested by two different approaches: gradients of primer annealing temperatures and gradients of viral concentration in samples. In the first approach, we carried out three multiplex PCRs, one for each virus, using annealing temperatures between 48 °C to 65 °C. All target regions of the HIV-1 multiplex PCR yielded positive amplifications at a range of temperatures from 48 °C to 58.3 °C. The amplification of the HIV1_108 and HIV1_222 targets started to disappear at temperatures above 58.3 °C (Fig. 11A). It is further noted that nonspecific amplifications were not observed at low temperatures (48 °C). The HCV multiplex PCR amplified well at temperatures between 54.5 °C and 58.3 °C. The amplification of some HCV target regions lost quality at temperatures below 54.5 °C. On the other hand, the smaller amplicons (HCV_88 and HCV_124) were not amplified at temperatures above 58.3 °C (Fig. 11B). Finally, the HBV multiplex PCR amplified better at temperatures below 54.5 °C, although all markers were amplified in the range of tested temperatures (Fig. 11C).

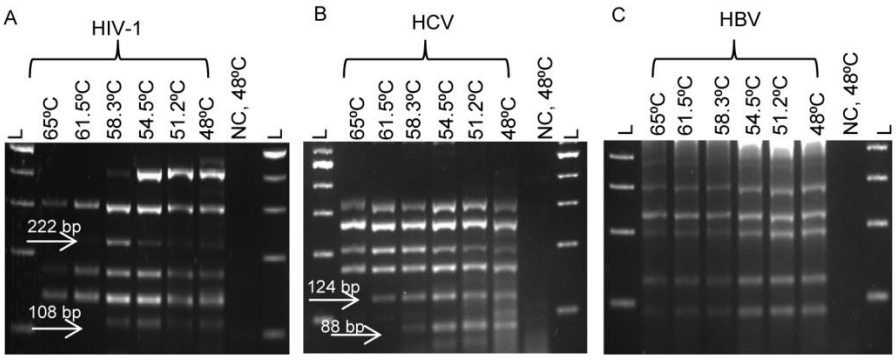


Fig. 8: Sensitivity test of multiplex PCR using gradients of annealing temperatures between 48 °C and 65 °C for HIV-1 (A), HCV (B) and HBV (C). One sample of each virus (1IV135, CV1A-6 and BV151) was used. NC - negative control (at the lowest temperature); L - 100 bp DNA ladder**.**

The second approach to test the sensitivity of the multiplex PCRs was based on a serial dilution of viral concentrations. The amplification of the diluted samples with the multiplex PCR revealed that the detection limit for HIV-1 was between 687.5 copies/mL (1:64) and 343.75 copies/mL (1:128). The markers HIV1_133, HIV1_300 and HIV1_222 were lost at concentrations equal or lower than 2,750 copies/mL, dilution 1:16 (Fig. 12A). In the case of HCV, loss of amplifications was observed at concentrations between 837.13 IU/mL (1:128) and 418.56 IU/mL (dilution 1:256), with good amplifications observed until the sample concentration of 1674.25 IU/mL, dilution 1:64 (Fig. 12B). The HBV serial dilution showed that amplifications failed at

concentrations between 39.84 IU/mL (1:2048) and 19.92 IU/mL (1:4096). However, no amplification was observed in markers HBV_89 and HBV_122 at concentrations below 19.92 IU/mL, dilution 1:4096 (Fig. 12C).

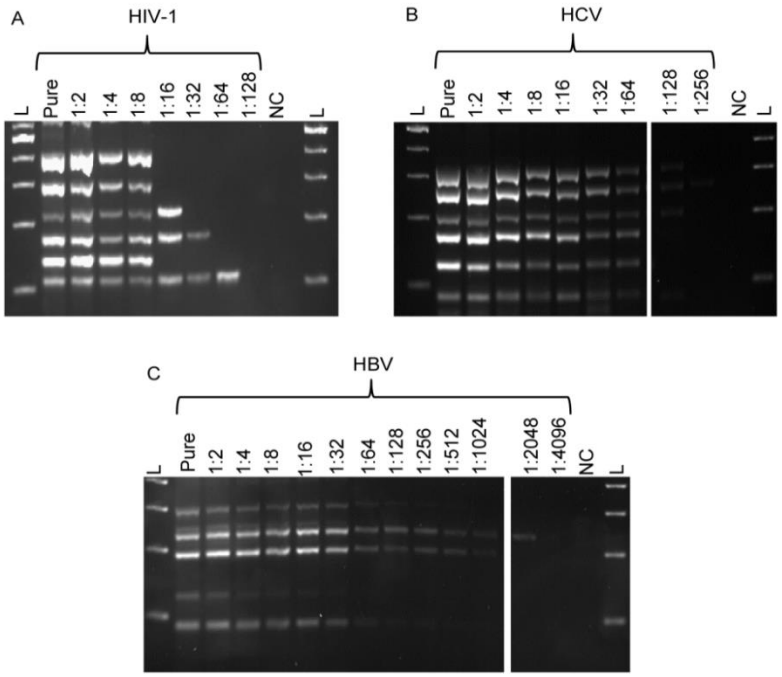


Fig. 9: Sensitivity test of multiplex PCRs using gradients of viral concentration for HIV-1 (A), HCV (B) and HBV (C). The starting point of concentration for HIV-1 sample 1IV001 was 44,000 copies/mL, for HCV sample CV1A-6 was 107,152 IU/mL and for HBV sample BV151 was 81,600 IU/mL. NC - negative control); L - 100 bp DNA ladder.

Finally, we tested the specificity of the three multiplex PCRs by using samples of virus not targeted by each one of the assays. For example, we tested the HIV-1 multiplex against HCV and HBV samples. We also tested the assays using human DNA from an uninfected individual. Our results demonstrate that the three multiplex PCRs are specific regarding the tested viral samples, since no nonspecific amplifications were observed, including in human DNA (Fig. 13A and 13B).

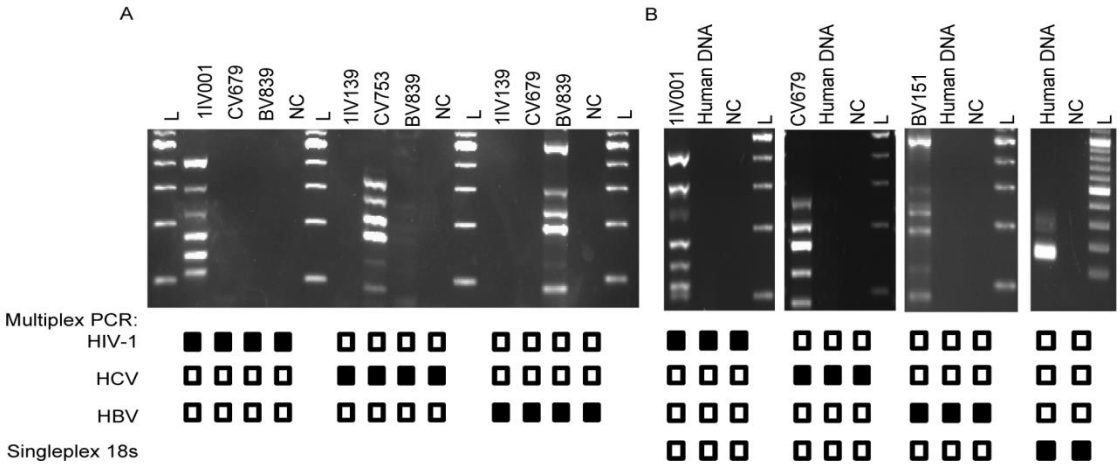


Fig. 10 (previous page): Specificity tests for the HIV-1, HCV and HBV multiplex PCRs. The agarose gels show the result of the test of the multiplex PCR for HIV-1, HCV and HBV using (A) samples of the three viruses and (B) human DNA. The 18S gene was used as positive control for presence of human DNA. NC - negative control); L - 100 bp DNA ladder.

Multiplex PCR for simultaneous detection of HIV-1, HCV and HBV

We selected six target regions (two for each virus) among the 16 target regions initially tested for all viruses (Fig. 4). The regions were chosen by having different lengths so the different viruses can be distinguished. The selected markers were HIV1_108 and HIV1_169 for HIV-1, HCV_206 and HCV_320 for HCV and HBV_89 and HBV_231 for HBV (Fig. 14). We also selected the regions with better amplification efficiencies in the individual multiplex PCRs of each virus (Fig. 6A, 8A and 10A).

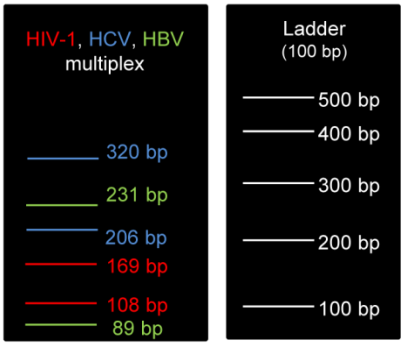


Fig. 11: Expected band pattern of the multiplex PCR for simultaneous detection of HIV-1, HCV and HBV. The multiplex PCR includes two specific target regions from each virus (expected amplicons for HIV-1 indicated in red, for HCV in blue and for HBV in green). The image on the right shows a 100 bp ladder.

each sample were correctly identified in 83.3% (5 out of 6) of the co-infections. One sample with a HIV-1/HCV co-infection (CV1IV91) had the four expected amplicons: the two HIV-1 amplicons

(108 bp and 169 bp) and the two HCV amplicons (206 bp and 320 bp). The CV1IV97 sample with HIV-1/HCV co-infection only amplified two target regions (one for each virus, HIV1_108 and HCV_206), allowing the detection of both viruses. Unfortunately, the sample CV1IV13 only amplified the two HCV markers with an expected length of 206 bp and 320 bp. The absence of amplification of the HIV-1 regions was probably due to the low HIV-1 viral load (41 copies/mL), which was much lower than the estimated for HCV (3,180,000 IU/mL). The BV1IV45 sample had the four target regions amplified, namely the two HIV-1 amplicons (108 bp and 169 bp) and two HBV amplicons (89 bp and 231 bp), as shown in Fig.15A.

The two artificial co-infections typed with the multiplex PCR showed amplifications for all target regions (Fig. 15B), suggesting that it can work in the presence of genetic material from the three target viruses.

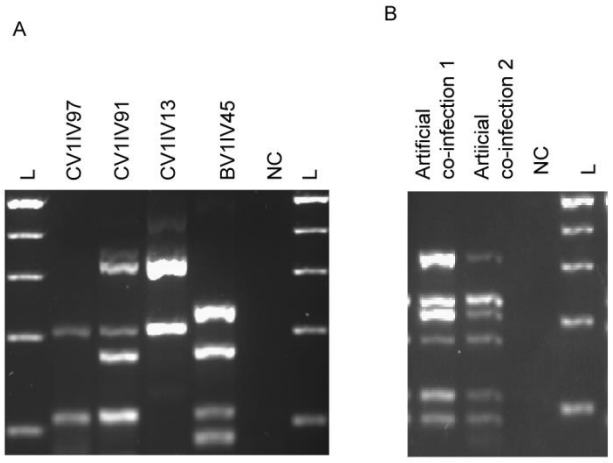


Fig. 12: Multiplex PCR for simultaneous detection of HIV-1, HCV and HBV (A) Agarose gels showing the amplifications with the multiplex PCR of (A) 3 HIV-1/HCV and 1 HIV-1/HBV co-infections and (B) 2 artificial co-infections with the three viruses. NC - negative control; L - 100 bp DNA ladder.

Discussion

The use of accurate methods for clinical diagnosis is an important aspect to control viral infections. Unfortunately, some of the available methods lack detection accuracy and are expensive and time-consuming. It is therefore of paramount importance to develop methods for the simultaneous identification of different viruses at low cost and with high efficiency levels. In this work, a multiplex PCR assay for simultaneous identification of three blood-borne viruses (HIV-1, HCV and HBV) and three individual multiplex PCRs for identification of each virus were developed.

One advantage of our methods is the placement of PCR primers on the most conserved viral genomic regions, allowing the detection of different virus genotypes. At least one region was amplified in all tested samples (n=70), while 26 samples had amplifications in all targeted regions (31.3 % of HIV-1, 50 % of HCV and 33.3 % of HBV samples). It should be noted that the HCV multiplex had the best amplification performance despite having six target regions on the multiplex PCR. The low percentage of HIV-1 samples that amplify for all markers might result from the high mutation rate and extreme genetic diversity among HIV-1 subtypes, which differ from each other by ~30 % over the complete genome [55]. The low percentage may also result from the low viral loads of HIV-1 samples when compared with those from HCV (Table S1). The HBV has low genetic diversity values, with only 8 to 15 % of divergence between the major genotypes [56]. However, the success of amplification of all HBV markers (33.3 %) was similar to the observed in HIV-1 (3.31 %). This result can be influenced by the low HBV samples size and by the lack of amplification observed in marker HBV_122, which only amplified in three HBV samples. The multiplex PCR for detection of the three viruses identified correctly all viruses present in 83.3% of the tested co-infections.

Because the genome of the three viruses is relatively short, PCR amplifications between primers designed for different markers are inevitable if several markers are included in a multiplex PCR. In order to avoid this problem, we have selected markers with interactions that result in large amplicons that do not overcome the desired target regions. Only one of the predicted extra amplification (EX_HIV1_410) was observed in the HIV-1 multiplex PCR (Fig. 5) and two predicted extra amplifications (EX_HBV_428 and EX_HBV_517) were observed in the HBV multiplex PCR (Fig. 9). In fact, such extra amplifications can add additional information to the multiplex PCR assay. Nevertheless, such extra amplifications can compete with the desired target regions in the PCR and deplete some of the PCR reagents. However, we did not notice a significant decrease of the PCR amplifications due to this factor. In any case, we have avoided

these interactions in the multiplex PCR for the simultaneous identification of the three viruses, since only two markers from each one are being used. No extra amplifications were observed in this multiple PCR (Fig. 15).

We have also optimized our assays taking into account some PCR conditions, like the melting temperature and the number of cycles. The gradients of annealing temperatures showed that robust amplifications can be obtained between 54.5 °C to 58 °C for HIV-1, HCV and HBV. However, the multiplex PCR performed very well at lower or higher temperatures. For instance, the HBV multiplex PCR amplified all target regions without non-specific amplifications at 51.2 °C and 48 °C. These results suggest that most markers included in our multiplex PCRs can be used in very different conditions, such as in real-time PCR or combined with other markers in different multiplex PCRs.

The use of short target regions is usually an advantage in the analysis of low quantity/quality template. For instance, HIV1_108, HIV_133 and HIV1_169 were the markers with the highest amplification success. Nevertheless, better amplifications were observed for the largest HCV and HBV target regions (Figs. 8A/B and 10A/B).

Our screening of samples from infected individuals allowed us to evaluate the amplification success of the selected target regions. Despite our efforts to design the best possible PCR primers in such highly variable genomes, some target regions were not amplified in some samples (Fig. 6A/B, 8A/B, 10A/B). The absence of amplification could be related with the level of conservation in the primer binding sites. It is well known that polymorphisms in primer-binding sites can lead to false-negative results, which are likely in HIV-1 and HCV where an extreme genetic diversity prevail [57].

In order to test such hypothesis, the conservation values for each primer (Table S2) and the success of PCR amplifications were compared for each target region (Fig. 6A, 8A, 10A). However, we found that HIV-1 and HBV markers with the highest amplification efficiency (HIV1_108 and HBV_231) are those with more polymorphic positions, on average (Table S2). The HCV_206 and HCV_320 markers had the highest success of amplification (92.31 % and 88.46 % of samples, respectively) as well as less polymorphic positions on average than the other markers. The exception is HCV_169 which had the lowest number of polymorphic positions and was amplified in 69.23 % of samples. In any case, the localization of HCV_169 in the genomic region less represented in the sequence alignment (5'UTR) may explain the low number of polymorphisms.

The different efficiencies in PCR amplifications observed among marker may result from other factors inherent to the PCR. The failure to amplify some HIV-1 target regions in samples with low viral load (which required a second PCR) suggests that the template concentration may explain the absence of amplification in some samples (Fig. 6F, 8F, 10F). In any case, all HIV-1 samples amplified at least one target region in the first PCR. Other factors may also influence our results, such as template nucleotide composition, formation of secondary structures in the template or primers, primer interactions, etc. Further validation studies are necessary to identify the causes of the missing amplifications.

One of the most notorious advantages of our method is the use of several target regions, which presents a clear advantage over methods targeting a single region. In cases where one or more target regions fail to amplify by the reasons described above or others, a correct identification is still possible based on the information from the remaining regions. For example, the presence of HIV-1 was detected in 1IVB8 sample with only a single amplified region of the multiplex PCR. Therefore, the occurrence of variability in primer-binding sites and other causes of missing PCR amplifications do not pose serious problems to our approach by the use of multiple target regions (Fig. 4). Our screening of clinical samples revealed that at least one of the markers was amplified in all cases (Fig. 5, 7, 9).

Other available methods either target a single genomic region (prone to false-negative results) or require multiple reactions, making the method expensive and time-consuming [29]. For example, real-time PCR assays using nonspecific DNA binding probes (e.g., SYBR Green I) are limited to the detection of the total amount of amplification products, which is prone to false-positives by non-specific amplifications or false-negatives by targeting usually a single genomic region [31]. The real-time PCR assays using viral specific probes (e.g. TaqMan probes) are expensive and limited in the number of probes that can be used simultaneously [31]. The AMPLINAT MPX test is an example of such type of assays [57]. This TaqMan real-time PCR is unable to discriminate among the three viruses selected in this work, requiring a secondary diagnostic step using virus-specific amplifications. The COBAS TaqMan HIV-1 or HCV (Roche Molecular Systems, Inc.) is based on singleplex PCR and, thus, prone to false-negative results.

The lack of sensitivity in NAT assays is aggravated when using a pool of mixed samples to reduce costs and labour in some screening plans [30]. Our multiplex PCR for simultaneous identification of the three viruses may reduce the need of using pools of samples since the detection of these viruses is possible in a single reaction.

A landmark in clinical diagnosis was the implementation of serological tests and NAT assays for detection of HIV-1, HCV and HBV, which reduced significantly the risk of viral transmission by blood donations. Nevertheless, a residual transmission risk remains to be eliminated [58]. The absence of antiviral antibodies to be detected by serological tests remains the main cause of the residual risk of viral transmission in blood transfusions [20]. Early in HIV-1 infections, prior to p24 antigen positivity, HIV-1 viral load range from $2x10^2$ to $1x10^6$ copies/mL. Studies report that a titer of HCV RNA can be very high in early infection, i.e., HCV viral load can reach $1x10^5$ – $1x10^6$ IU/mL and this value is maintained until antibody detection [58]. During HBV seroconversion window period, HBsAg is the earliest detectable serologic marker [14]. However, HBsAg is often undetectable and serological tests result in false negatives. For example, AMPLINAT MPX system detects HBV DNA in samples that were negative for HBsAg test. In the window period, HBV viral load is ranging between 10 – $1x10^4$ IU/mL [57, 59]. By these reason, knowing the detection limit of a method is of great relevance.

In order to assess the sensitivity of our assays, one plasma sample of each virus was serial diluted to determine detection limits. The limits determined for the multiplex PCR were 343.75 - 687.5 copies/mL for HIV-1, 418 - 837.13 IU/mL for HCV and 19 - 39.84 IU/mL for HBV. These values suggest that our multiplex PCRs should be sufficient to detect viral infections during the window period. In any case, HIV-1 samples with viral loads lower than 343.75 copies/mL and HBV samples with viral loads lower than 19 IU/mL may not be correctly diagnosed by this method with the conditions implemented in the PCR (e.g., 40 cycles for HIV-1 and 35 cycles for HBV). In these cases, a second PCR with more cycles using as template the PCR products from a first PCR could be used to guarantee the absence of viral genetic material. We used a second PCR for a better amplification of some markers in 11 HIV-1 samples with low viral loads, although the first PCR was always enough for HIV-1 detection. Other techniques face the same problem when screening HIV-1 samples with viral loads. For instance, 46.8 PCR cycles were necessary in the real time PCR array described by Pripuzova et al [26] to detect HIV-1 in a sample whit a viral load of 51 copies/mL.

It should be noted that the viral concentrations on the plasma samples used to calculate the detection limits were estimated using a real-time PCR performed on the laboratories providing the samples. Although it is a good approximation to real values, it is possible that our RNA isolation and cDNA synthesis (HIV-1 and HCV) or DNA extraction (HBV) would lead to different viral concentrations in the template samples used in the serial dilution PCR. Future sensibility tests using standard reference samples will be used to confirm the multiplex PCR detection

limits. The detection limits of the multiplex PCR for the simultaneous identification of the three viruses should not differ significantly from the estimated for the individual multiplex PCR, since the same target regions are being used. However, a sensitivity test for this multiplex PCR is recommended, considering that interactions in the new mixture of primers and other PCR factors may affect the amplification efficiency.

We also tested the specificity of our multiplex PCR assays using genetic material from human and non-target viruses. The assays proved to be 100 % specific for the targeted virus, since no unspecific amplifications were detected. By this reason, false positive results are unlike using our methods.

In conclusion, we confirmed previous studies that demonstrate the advantages of using multiplex PCR for viral identifications in various clinical and epidemiological settings [38]. In this thesis, a conventional multiplex PCR was developed for simultaneously detection of the major blood-borne viruses, HIV1-, HCV and HBV. The simplicity of the assays and the fact that the detection is based on the amplification of several viral genomic regions are the two main advantages. The identification can be obtained using conventional agarose or acrylamide gels, with no need for expensive sequencing or real-time PCR apparatus. These multiplex PCR assays are sensitive and specific, although more tests are required (e.g., standards samples) to obtain greater precision in sensitivity and specificity. The direct detection of multiple viruses in a single reaction is an improvement for the screening of samples in the clinical context or in blood banks. Our methods can also be used in early infant diagnosis, where serological tests have a high rate of false positives due to the detection of maternal antiviral antibodies in the infants [53]. Our multiplex PCRs can serve as a complement to techniques routinely used in laboratories or as a screening technique prior to other analyses, by reducing costs and increasing the accuracy of the detection.

References

1. Cann, A., *Principles of molecular virology*. Vol. 1. 2001: Academic Press.
2. Haaheim, L.R., J.R. Pattison, and R.J. Whitley, *A practical guide to clinical virology*. 2002: John Wiley & Sons.
3. WHO. *Global summary of the AIDS epidemic* 2013 10 September 2014]; Available from: http://www.who.int/hiv/data/epi_core_dec2014.png?ua=1.
4. WHO. *Hepatitis C fact sheet*. 2014 10 September 2014]; Available from: http://www.who.int/mediacentre/factsheets/fs164/en/.
5. WHO. *Hepatitis B fact sheet* 2014 10 September 2014]; Available from: http://www.who.int/mediacentre/factsheets/fs204/en/.
6. WHO. *World Hepatitis day 2012*. 2012; Available from: http://www.who.int/csr/disease/hepatitis/world_hepatitis_day/en/.
7. Alter, M.J., *Epidemiology of viral hepatitis and HIV co-infection.* Journal of hepatology, 2006. **44**: p. S6-S9.
8. Chu, C.J. and S.D. Lee, *Hepatitis B virus/hepatitis C virus coinfection: epidemiology, clinical features, viral interactions and treatment.* Journal of gastroenterology and hepatology, 2008. **23**(4): p. 512-520.
9. Lincoln, D., K. Petoumenos, and G. Dore, *HIV/HBV and HIV/HCV coinfection, and outcomes following highly active antiretroviral therapy.* HIV medicine, 2003. **4**(3): p. 241-249.
10. Maier, I. and G.Y. Wu, *Hepatitis C and HIV co-infection: a review.* World Journal of Gastroenterology, 2002. **8**(4): p. 577-579.
11. Hulo, C., et al., *ViralZone: a knowledge resource to understand virus diversity.* Nucleic acids research, 2011. **39**(suppl 1): p. D576-D582.
12. Michael, N., *HIV Protocols*. 1999: Humana Press.
13. Tang, H., *Hepatitis C: Methods and Protocols*. 2008: Humana Press.
14. Hamatake, R.K. and J.Y.N. Lau, *Hepatitis B and D Protocols: Volume 1: Detection, Genotypes, and Characterization*. 2004: Humana Press.
15. Sayre, K.R., et al., *False-positive human immunodeficiency virus type 1 western blot tests in noninfected blood donors.* Transfusion, 1996. **36**(1): p. 45-52.
16. Widell, A., et al., *Transmission of Hepatitis C via both Erythrocyte and Platelet Transfusions from a Single Donor in Serological Window-Phase of Hepatitis C.* Vox sanguinis, 1996. **71**(1): p. 55-57.
17. Candotti, D., et al., *AIDS in an HIV-seronegative Ghanaian woman with intersubtype A/G recombinant HIV-1 infection.* Journal of medical virology, 2000. **62**(1): p. 1-8.
18. Allain, J.P., et al., *Evidence that anti-HBc but not HBV DNA testing may prevent some HBV transmission by transfusion.* British journal of haematology, 1999. **107**(1): p. 186-195.
19. Busch, M.P., S.H. Kleinman, and G.J. Nemo, *Current and emerging infectious risks of blood transfusions.* Jama, 2003. **289**(8): p. 959-962.
20. Busch, M., et al., *Nucleic acid amplification testing of blood donors for transfusion-transmitted infectious diseases.* Transfusion, 2000. **40**(2): p. 143-159.
21. Yang, S. and R.E. Rothman, *PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings.* The Lancet infectious diseases, 2004. **4**(6): p. 337-348.
22. Tang, Y.-W., G.W. Procop, and D.H. Persing, *Molecular diagnostics of infectious diseases.* Clinical Chemistry, 1997. **43**(11): p. 2021-2038.
23. Niesters, H., *Molecular and diagnostic clinical virology in real time.* Clinical microbiology and infection, 2004. **10**(1): p. 5-11.
24. Fox, J.D., *Nucleic acid amplification tests for detection of respiratory viruses.* Journal of Clinical Virology, 2007. **40**: p. S15-S23.

25. Yu, D., et al., *Rapid detection of common viruses using multi-analyte suspension arrays.* Journal of virological methods, 2011. **177**(1): p. 64-70.
26. Pripuzova, N., et al., *Development of Real-Time PCR Array for Simultaneous Detection of Eight Human Blood-Borne Viral Pathogens.* PloS one, 2012. **7**(8): p. e43246.
27. Jackson, B., et al., *The cost-effectiveness of NAT for HIV, HCV, and HBV in whole-blood donations.* Transfusion, 2003. **43**(6): p. 721-729.
28. Allain, J.P., *Genomic screening for blood-borne viruses in transfusion settings.* Clinical & Laboratory Haematology, 2000. **22**(1): p. 1-10.
29. Vernet, G., *Molecular diagnostics in virology.* Journal of clinical virology, 2004. **31**(4): p. 239-247.
30. Stramer, S., et al. *Comparison of HIV-1/HCV nucleic acid test reactivities of diluted and undiluted donations*. in *Transfusion*. 2000. AMER ASSOC BLOOD BANKS 8101 GLENBROOK RD, BETHESDA, MD 20814-2749 USA.
31. Mackay, I.M., K.E. Arden, and A. Nitsche, *Real-time PCR in virology.* Nucleic acids research, 2002. **30**(6): p. 1292-1305.
32. Gibellini, D., et al., *Simultaneous detection of HCV and HIV-1 by SYBR Green real time multiplex RT-PCR technique in plasma samples.* Molecular and cellular probes, 2006. **20**(3): p. 223-229.
33. Candotti, D., et al., *Multiplex real-time quantitative RT-PCR assay for hepatitis B virus, hepatitis C virus, and human immunodeficiency virus type 1.* Journal of virological methods, 2004. **118**(1): p. 39-47.
34. Dineva, M.A., et al., *Simultaneous visual detection of multiple viral amplicons by dipstick assay.* Journal of clinical microbiology, 2005. **43**(8): p. 4015-4021.
35. Adami, V., et al., *Qualitative multiplex RT-PCR for simultaneous detection of hepatitis C virus and human immunodeficiency virus in plasma samples.* Clinical microbiology and infection, 2004. **10**(12): p. 1075-1080.
36. Heredia, A., et al., *Development of a multiplex PCR assay for the simultaneous detection and discrimination of HIV-1, HIV-2, HTLV-I and HTLV-II.* Clinical and diagnostic virology, 1996. **7**(2): p. 85-92.
37. Kirschberg, O., et al., *A multiplex-PCR to identify hepatitis B virus—genotypes A–F.* Journal of clinical virology, 2004. **29**(1): p. 39-43.
38. Elnifro, E.M., et al., *Multiplex PCR: optimization and application in diagnostic virology.* Clinical microbiology reviews, 2000. **13**(4): p. 559-570.
39. Edwards, M.C. and R.A. Gibbs, *Multiplex PCR: advantages, development, and applications.* Genome Research, 1994. **3**(4): p. S65-S75.
40. Markoulatos, P., N. Siafakas, and M. Moncany, *Multiplex polymerase chain reaction: a practical approach.* Journal of clinical laboratory analysis, 2002. **16**(1): p. 47-51.
41. Kuiken, C., B. Korber, and R.W. Shafer, *HIV sequence databases.* AIDS reviews, 2003. **5**(1): p. 52.
42. Kuiken, C., et al., *HIV sequence compendium 2012.* Los Alamos, New Mexico: Los Alamos National Laboratory, Theoretical Biology and Biophysics, 2012.
43. Finn, R.D., J. Clements, and S.R. Eddy, *HMMER web server: interactive sequence similarity searching.* Nucleic acids research, 2011: p. gkr367.
44. Kuiken, C., et al., *The Los Alamos hepatitis C sequence database.* Bioinformatics, 2005. **21**(3): p. 379-384.
45. Pruitt, K.D., T. Tatusova, and D.R. Maglott, *NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.* Nucleic acids research, 2007. **35**(suppl 1): p. D61-D65.
46. Harrison, A., et al., *Genomic analysis of hepatitis B virus reveals antigen state and genotype as sources of evolutionary rate variation.* Viruses, 2011. **3**(2): p. 83-101.

47. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic acids research, 2004. **32**(5): p. 1792-1797.
48. Carneiro, J., A. Resende, and F. Pereira. *IGVID: A database of informative genomic regions for virus identification*. 2014 10 September 2014]; Available from: http://igvid.portugene.com.
49. Kibbe, W.A., *OligoCalc: an online oligonucleotide properties calculator.* Nucleic acids research, 2007. **35**(suppl 2): p. W43-W46.
50. Vallone, P.M. and J.M. Butler, *AutoDimer: a screening tool for primer-dimer and hairpin structures.* Biotechniques, 2004. **37**(2): p. 226-231.
51. NCBI. *Basic Local Alignment Search Tool (BLAST)*. 2009 29 August 2014]; Available from: http://blast.ncbi.nlm.nih.gov/Blast.cgi.
52. Drummond, A., et al., *Geneious v5. 4*, 2011.
53. Boyle, D.S., et al., *Rapid detection of HIV-1 proviral DNA for early infant diagnosis using recombinase polymerase amplification.* MBio, 2013. **4**(2): p. e00135-13.
54. Cleland, A., et al., *Development of multiplexed nucleic acid testing for human immunodeficiency virus type 1 and hepatitis C virus.* Vox sanguinis, 2001. **81**(2): p. 93-101.
55. Janssens, W., A. Buvé, and J.N. Nkengasong, *The puzzle of HIV-1 subtypes in Africa.* Aids, 1997. **11**(6): p. 705-712.
56. Kidd-Ljunggren, K., Y. Miyakawa, and A.H. Kidd, *Genetic variability in hepatitis B viruses.* Journal of General Virology, 2002. **83**(6): p. 1267-1280.
57. Meng, Q., et al., *Automated multiplex assay system for simultaneous detection of hepatitis B virus DNA, hepatitis C virus RNA, and human immunodeficiency virus type 1 RNA.* Journal of clinical microbiology, 2001. **39**(8): p. 2937-2945.
58. Stramer, S.L., et al., *Detection of HIV-1 and HCV infections among antibody-negative blood donors by nucleic acid–amplification testing.* New England Journal of Medicine, 2004. **351**(8): p. 760-768.
59. Biswas, R., et al., *Comparative sensitivity of HBV NATs and HBsAg assays for detection of acute HBV infection.* Transfusion, 2003. **43**(6): p. 788-798.

Supplementary material

Supplementary figures

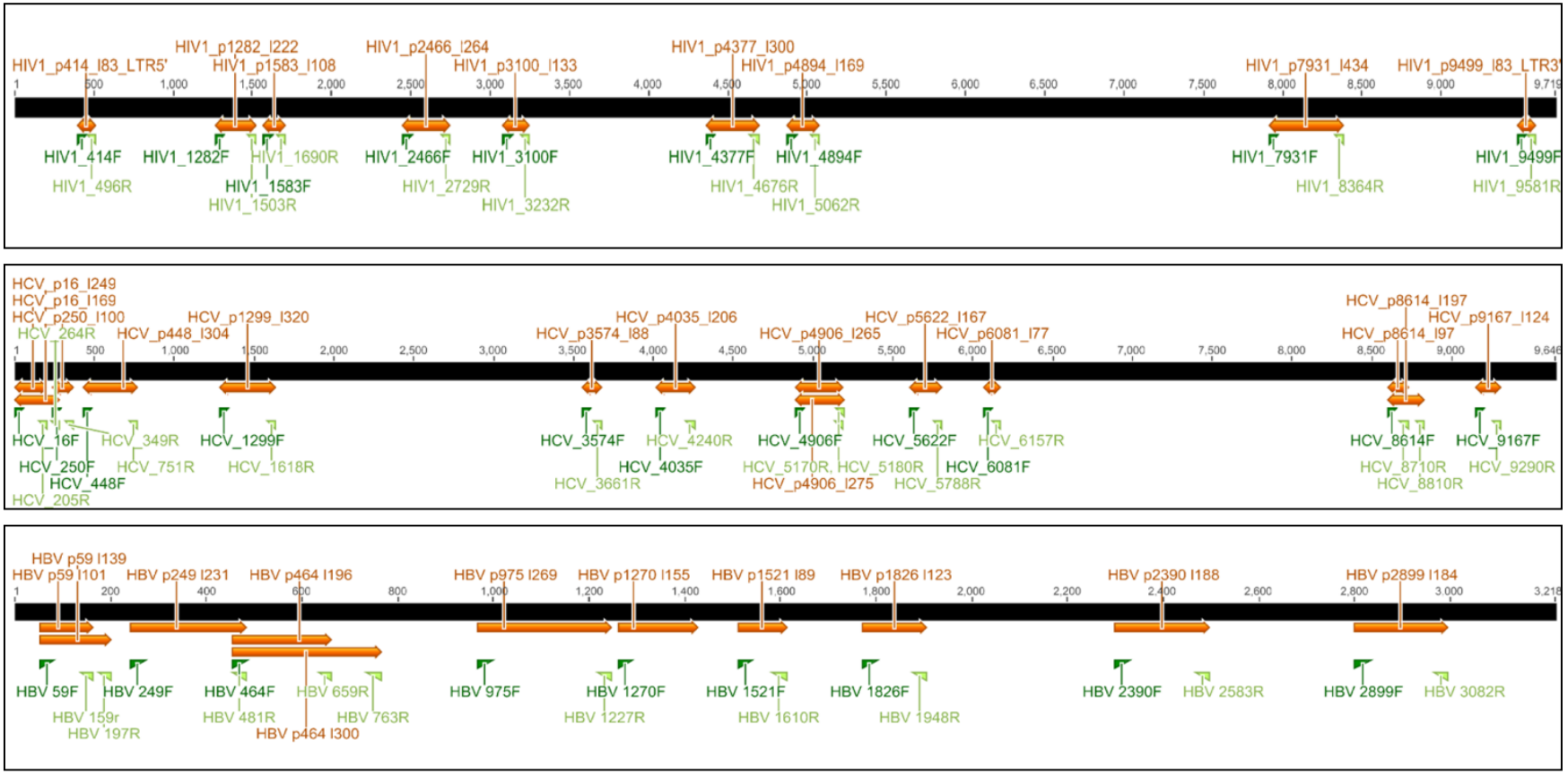


Fig. S1: Schematic representation of the localization of target regions (orange arrows) and PCR primers (green arrows) in HIV-1, HCV and HBV references genome sequences.

Supplementary tables

Table S1: Description of the plasma samples used in this study.

Virus	Lab code	Genotype	Viral load
HIV-1	1IV001	G	44000
	1IV002	A1	98500
	1IV003	nd	15400
	1IV400	G	43500
	1IV127	A1	10900
	1IV139	B	98200
	1IV268	B	56900
	1IV135	nd	412000
	1IVB1	B	360
	1IVB2	B	2803
	1IVB3	B	544436
	1IVB4	B	168891
	1IVB5	B	20123
	1IVB6	B	4130
	1IVB7	B	750
	1IVB8	B	118710
	1IVB10	B	953
	1IVC1	C	16754
	1IVC2	C	4702
	1IVC3	C	2214
	1IVC4	C	824
	1IVC5	C	50998
	1IVC6	C	111
	1IVC7	C	759
	1IVC8	C	702
	1IVC9	C	Unknown
	1IVC10	C	18728
	1IVCP1	Complex	1044
	1IVCP2	Complex	Unknown
	1IVCP3	Complex	25510
	1IVCP4	Complex	56872
	1IVCP5	Complex	585
HCV	CV001	nd	2450000
	CV002	nd	9790000
	CV003	3A	11000000
HCV	CV641	3A	4700000
	CV683	4A/4C/4D	2670
	CV679	3A	423000
	CV753	nd	1510000
	CV3-1	3	67608.30
	CV3-2	3	19054.60
	CV3-3	3	4466.84
	CV3-4	3	562341
	CV3-5	3	2344229
	CV3-6	3	758578
	CV3-7	3	10232929.92
	CV3-8	3	34673.69
	CV3-9	3	501187
	CV3-10	3	1348962.90
	CV1	1A	52480.70
	CV2	1A	181970
	CV3	1A	5623413.25
	CV4	1A	13803.8
	CV5	1A	1318257
	CV1A-6	1A	107152
	CV1A-7	1A	3981072
	CV1A-8	1A	1584893
	CV1A-9	1A	7943282
HBV	BV269	nd	237
	BV423	nd	20
	BV609	D	22600
	BV646	D	3050
	BV839	D	89500
	BV013	A	170000000
	BV151	nd	81600
	BV117	nd	20
Co-infections	CV1IV97	1A (HCV), nd (HIV-1)	1440000 / 2650
	CV1IV13	nd (HCV), nd (HIV-1)	3180000 / 41
	CV1IV91	3A (HCV), nd (HIV-1)	1880000 / 110000
	BV1IV45	A (HBV), nd (HIV-1)	170000000 / 43700

The table describes the laboratory code, genotype and viral load of samples from HIV-1 (n=32), HCV (n=26) and HBV (n=8) infected individuals, as well as 4 samples from co-infected individuals (1 with a HIV-1/HBV co-infection and 3 with HIV-1/HCV co-infections). The units of the HIV-1 viral load is in copies/mL and of the HCV and HBV in IU/mL.

Table S2: Description of the target regions and PCR primers used in the multiplex PCRs for identification of HIV-1, HCV and HBV

Target region					Primer						
Virus	Name	Nº sequences in alignment	Genomic region	Length (mean and std dev)	Name	Sequence (5' – 3')	Forward/ reverse	Tm (°C)	Length (nt)	Nº of variable positions (nt)	Nº of variable positions at 3' end (nt)
HIV-1	HIV1_LTR	31	LTR	82.7 ± 0.4	HIV1_414F	CTCAGATCCTGCATATAAGCAG	FW	60.1	22	7	0
					HIV1_496R	CAGAGAGCTCCCAGGCTC	RV	60.8	18	4	0
	HIV1_222*	170	*gag*	222 ± 0.0	HIV1_1282F	AGCCCAGAAGTGATACCCATG	FW	61.2	21	8	1
					HIV1_1507R	TCCTGCTATGTCACTTCCCC	RV	60.5	20	10	3
	HIV1_108*	170	*gag*	108 ± 0.0	HIV1_1583F	GGATAATCCTGGGATTAAATAAAATA	FW	58.4	26	11	1
					HIV1_1690R	AGAACCGGTCTACATAGTCTCT	RV	60.1	22	10	2
	HIV1_264	170	*pol*	264 ± 0.2	HIV1_2466F	ATAGGTAC GTATTAGTAGGACCT	FW	60.3	24	15	2
					HIV1_2729R	GAAAATCCATACAATACTCCAGTA	RV	58.3	24	11	1
	HIV1_133*	170	*pol*	133.0 ± 0.1	HIV1_3100F	TGGATGATTTGTATGTAGGATCTGA	FW	60.9	25	10	1
					HIV1_3232R	AGGAATGGAGGTTCTTTCTGATG	RV	60.9	23	9	1
	HIV1_300*	170	*pol*	300.0 ± 0.0	HIV1_4377F	ATGCATGGACAAGTAGACTGTAG	FW	60.9	23	11	2
					HIV1_4676R	TCCTTGACTTTGGGGATTGTAG	RV	60.1	22	8	2
	HIV1_169*	170	*pol*	169.0 ± 0.4	HIV1_4894F	ATTTTCGGGTTTATTACAGGGACA	FW	60.3	24	7	1
					HIV1_5062R	TCACCTGCCATCTGTTTTCCA	RV	59.5	21	9	2
	HIV1_434	170	*tat*. *env*	434.1 ± 0.4	HIV1_7931F	AGTCTGGGGCATCAAGCAG	FW	59.5	19	11	2
					HIV1_8364R	GTGAATATCCCTGCCTAACTC	RV	59.5	21	13	5
HCV	HCV_169*	140	5'UTR	168.8 ± 0.4	HCV_16F	GGGGCGACACTCCACCA	FW	59.8	17	3	1
					HCV_205R	CGTCCTGGCAATTCCGGTG	RV	61.6	19	8	0
	HCV_249	140	5'UTR	248.9 ± 0.4	HCV_16F	GGGGCGACACTCCACCA	FW	59.8	17	3	1
					HCV_264R	ACACTACTCGGCTAGCAGTCT	RV	61.2	21	4	2
	HCV_100	467	5'UTR	100 ± 0.0	HCV_250F	CTAGCCGAGTAGTGTTGGGT	FW	60.5	20	3	1
					HCV_349R	GTGCTCATGGTGCACGGT	RV	58.4	18	5	3

Target region					Primer						
Virus	Name	Nº sequences in alignment	Genomic region	Length (mean and std dev)	Name	Sequence (5' – 3')	Forward/ reverse	Tm (°C)	Length (nt)	Nº of variable positions (nt)	Nº of variable positions at 3' end (nt)
HCV	HCV_304	471	Core	304 ± 0.0	HCV_448F	TGTTGCCGCGCAGGGGC	FW	61.8	17	8	2
					HCV_751R	ATGTACCCCATGAGGTCGG	RV	59.5	19	8	2
	HCV_320*	471	E1	319.7 ± 4.0	HCV_1299F	TGGGATATGATGATGAACTGG	FW	57.5	21	8	2
					HCV_1618R	GCCGTGCTATTGATGTGCCA	RV	60.5	20	10	1
	HCV_88*	471	NS3	88.0 ± 0.0	HCV_3574F	GCTGGACTGTCTACCACGG	FW	61.6	19	12	2
					HCV_3661R	TCTTGGTCCACATTGGTATACAT	RV	59.2	23	16	2
	HCV_206*	471	NS3	206.0 ± 0.1	HCV_4035F	ACCGGCAGCGGTAAGAGC	FW	60.8	18	9	3
					HCV_4240R	AGGAACTTGCCGTAGGTGGA	RV	60.5	20	11	2
	HCV_265*	471	NS3	265.0 ± 0.0	HCV_4906F	ACGCGGGCTGTGCTTGGT	FW	60.8	18	10	1
					HCV_5170R	CACTTCCACATCTGGTCCCA	RV	60.5	20	11	0
	HCV_275	471	NS3	275.0 ± 0.0	HCV_4906F	ACGCGGGCTGTGCTTGGT	FW	60.8	18	10	1
					HCV_5180R	GCGGATCAAACACTTCCACAT	RV	59.5	21	15	4
	HCV_167	471	NS4B	167.0 ± 0.0	HCV_5622F	TGGGCGAAGCACATGTGGAA	FW	60.5	20	11	1
					HCV_5788R	ACCCACCCCCCCAATATGTT	RV	59.5	21	15	4
	HCV_77	471	NS4B	77.0 ± 0.0	HCV_6081F	CAATGGATGAACCGGCTAATAGC	FW	62.9	23	8	1
					HCV_6157R	TCGCTCTCCGGCACGTAGT	RV	61.6	19	11	2
	HCV_97	471	NS5B	97.0 ± 0.0	HCV_8614F	CCTTCACGGAGGCTATGAC	FW	59.5	19	12	1
					HCV_8710R	ACGTTGGAGGAGCATGATGTTAT	RV	60.9	23	13	4
	HCV_197	471	NS5B	197.0 ± 0.0	HCV_8614F	CCTTCACGGAGGCTATGAC	FW	59.5	19	12	1
					HCV_8810R	AGTGTGTCTTGCTGTCTCCC	RV	60.5	20	16	2
	HCV_124*	471	NS5B	124.0 ± 0.1	HCV_9167F	CAAGTACCTCTTCAACTGGGC	FW	61.2	21	8	0
					HCV_9290R	GCTGTGATAAATGTCTCCCCC	RV	61.2	21	13	3

Target region					Primer						
Virus	Name	Nº sequences in alignment	Genomic region	Length (mean and std dev)	Name	Sequence (5' – 3')	Forward/ reverse	Tm (°C)	Length (nt)	Nº of variable positions (nt)	Nº of variable positions at 3' end (nt)
HBV	HBV_101	385	*P*, *S*	101.0 ± 0.0	HBV_56F	CCTGCTGGTGGCTCCAGT	FW	60.8	18	8	5
					HBV_156R	ATGTTCGGTGCAGGGTCC	RV	58.4	18	8	1
	HBV_139	386	*P*, *S*	138.9 ± 0.0	HBV_16F	CCTGCTGGTGGCTCCAGT	FW	60.8	18	8	5
					HBV_194R	CGAGCAGGGGTCCTAGGA	RV	60.8	18	5	2
	HBV_231*	386	*P*, *S*	231.0 ± 0.1	HBV_246F	GTCTAGACTCGTGGTGGAC	FW	59.5	19	7	2
					HBV_476R	GACAAACGGGCAACATACC	RV	57.5	19	7	1
	HBV_196	387	*P*, *S*	196.0 ± 0.5	HBV_459F	GTATGTTGCCCGTTTGTCCT	FW	58.4	20	7	1
					HBV_645R	GAGGCCCACTCCCATAGG	RV	60.8	18	3	0
	HBV_300*	387	*P*, *S*	300 ± 0.5	HBV_459F	GTATGTTGCCCGTTTGTCCT	FW	58.4	20	7	1
					HBV_758R	CCCAATACCACATCATCCATAT	RV	58.4	22	8	2
	HBV_269	387	*P*	268.6 ± 7.4	HBV_970F	CCTATTGATTGGAAAGTATGTCA	FW	57.6	23	10	3
					HBV_1238R	CGCATGCGCTGATGGCC	RV	59.8	17	8	2
	HBV_142	386	*P*, *X*	142.0 ± 0.0	HBV_1265F	TCCATACTGCGGAACTCCT	FW	57.5	19	7	3
					HBV_1406R	AGGATCCAGTTGGCAGCACA	RV	60.5	20	7	3
	HBV_89*	384	*P*, *X*	89.0 ± 0.0	HBV_1516F	ACCACGGGGCGCACCTC	FW	61.8	17	8	1
					HBV_1604R	ACGTGCAGAGGTGAAGCGA	RV	59.5	19	8	2
	HBV_122*	385	*X*, *C*	122.0 ± 0.0	HBV_1774F	TAGGAGGCTGTAGGCATAAATTG	FW	60.9	23	4	1
					HBV_1895R	AAAGCCACCCAAGGCACAG	RV	59.5	19	2	0
	HBV_188*	386	*P*, *C*	188.6 ± 1.8	HBV_2301F	CACCAAATGCCCCTATCTTATC	FW	60.1	22	7	2
					HBV_2494R	CCCGTAAAGTTTCCCACCTT	RV	58.4	20	5	0
	HBV_184	387	*P*, *S*	174.3±15.1	HBV_2809F	CCTCATTTTGTGGGTCACCATA	FW	60.1	22	10	1
					HBV_2992R	GGTTGAAGTCCCAATCTGGA	RV	58.4	20	8	2

A total of 33 target regions are described by name, number of sequences in alignment, corresponding genomic region and mean length with standard deviation in nucleotides (nt). A total of 66 primers (forward and reverse) are listed by name, primer sequence in 5'-3' orientation, melting temperature (°C), primer length (nt), number of polymorphic positions across alignments for the entire primer-binding region or for the last five bases at 3' end. * denotes the target regions included in individual multiplex PCRs.